IJCSIS Vol. 10 No. 8, August 2012
ISSN 1947-5500

International Journal of Computer Science & Information Security


Cornell University
Library


Cogprints


Google scholar


.docstoc
find and share professional documents


ScientificCommons

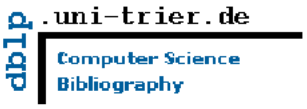

View my documents on
Scribd


BASE
Bielefeld Academic Search Engine


SCIRUS
search engine for science


SciRate.com


CiteSeerx beta


dblp .uni-trier.de
Computer Science
Bibliography


Q·Sensei


DOAJ
DIRECTORY OF
OPEN ACCESS
JOURNALS


EBSCO
HOST


ProQuest

Editorial
Message from Managing Editor

*The **International Journal of Computer Science and Information Security** (IJCSIS) focuses to publish the emerging area of computer applications and practices, and latest advances in cloud computing, information security, green IT etc. IJCSIS addresses innovative developments, research issues/solutions in computer science and related technologies. It is a well-established and notable venue for publishing high quality research papers as recognised by various universities, international professional bodies and Google scholar citations.*

IJCSIS editorial board solicits authors/researchers/scholars to contribute to the journal by submitting articles that illustrate research results, projects, surveying works and industrial experiences. The aim is also to allow academia promptly publish research work to sustain or further one's career.

For complete details about IJCSIS archives publications, abstracting/indexing, editorial board and other important information, please refer to IJCSIS homepage. IJCSIS appreciates all the insights and advice from authors/readers and reviewers. Indexed by the following International Agencies and institutions: Google Scholar, Bielefeld Academic Search Engine (BASE), CiteSeerX, SCIRUS, Cornell's University Library EI, Scopus, DBLP, DOI, ProQuest, EBSCO.

Google Scholar reported a large amount of cited papers published in IJCSIS. We will continue to encourage the readers, authors and reviewers and the computer science scientific community and authors to continue citing papers published by the journal. Considering the growing interest of academics worldwide to publish in IJCSIS, we invite universities and institutions to partner with us to further encourage open-access publications

*We look forward to receive your valuable papers. The topics covered by this journal are diverse. (See monthly Call for Papers). If you have further questions please do not hesitate to contact us at **ijcsiseditor@gmail.com**. Our team is committed to provide a quick and supportive service throughout the publication process.*

A complete list of journals can be found at:
http://sites.google.com/site/ijcsis/

IJCSIS Vol. 10, No. 8, August 2012 Edition

ISSN 1947-5500 © IJCSIS, USA.

Journal Indexed by (among others):

Google scholar ScientificCommons CiteSeerx beta SCIRUS search engine for science

DOAJ DIRECTORY OF OPEN ACCESS JOURNALS INDEX COPERNICUS INTERNATIONAL ProQuest Start here.

IJCSIS EDITORIAL BOARD

TABLE OF CONTENTS

A Novel Data Hiding Scheme for Binary Images

Do Van Tuan
Hanoi College of Commerce and Tourism
Hanoi – Vietnam
dvtuanest@gmail.com

Tran Dang Hien
Vietnam National University
hientd_68@yahoo.com

Pham Van At
Hanoi University of Communications and Transport
phamvanat83@vnn.vn

***Abstract* - this paper presents a new scheme for hiding a secret message in binary images. Given m×n cover image block, the new scheme can conceal as many as $\lfloor log_2(m \times n + 1) \rfloor$ bits of data in block, by changing at most one bit in the block. The hiding ability of the new scheme is the same as Chang et al.'s scheme and higher than Tseng et al.'s scheme. Additionally, the security of the new scheme is higher than the two above schemes.**

Keywords - Data hiding; steganography; security; binary image;

I. INTRODUCTION

Nowadays, the Internet is the most popular channel for data exchanges between providers and users. Yet, the data safety issue on the Internet is always a challenge to managers and researchers, as the data on the Internet is easily tampered with and stolen by hackers during transmission. In addition to encryption schemes, data hiding has an important role in secret message transmission, authentication, and copyright protection on public exchange environment.

Data hiding is a technique to conceal a secret message in cover media, to avoid arousing an attacker's attention. The cover media is often a document, image, audio or video. According to [1], the data hiding schemes proposed in an image can be divided into two categories. In the first category, the schemes hide a secret message in the spatial domain of the cover image [3,4,6,] and the least significant bits of each pixel in cover image is modified to hide the secret message. In the second category, the schemes hide a secret message in transformed domain of cover image [2,8]. Several transformation functions, such as discrete cosine transform and discrete wavelet transform are widely used.

However, most cover images of the above schemes are gray-level images or color images. The binary image is not often used in cover media [1,5,7]. The major reason is that the modification is easily detected when a single pixel is modified in a binary image. For binary images, two schemes are seen as modern and efficient in TCP scheme [5] and CTL scheme [1]. Accordingly, given an m×n cover image block from cover image, both schemes can conceal maximum $r = \lfloor log_2(m \times n + 1) \rfloor$ bits in block. To hide r bits, TCP scheme changes two pixels at most, but CTL scheme only need change one pixel at most. Therefore, the invisibility of CTL scheme is higher than TCP scheme. However, the content of the CTL scheme is quite complicated. This paper presents a novel scheme to hide a secret message in binary images. In addition, the hiding capacity and stego-image quality of new scheme are the same with CTL scheme, but the security property of the new scheme is higher than CTL scheme. Moreover, the content of new scheme is simpler than above two schemes.

The remaining text of this paper is organized as follows: In section 2, we define some operators used in this paper. In section 3, we present some hiding data algorithms in a block. These algorithms are background for new data hiding scheme presented in section 4. In section 5, we present some experimental results. Finally, Section 6 presents the conclusions.

II. NOTATION

Definition 1. Denote $\otimes$ is component-wise multiplication of two matrices of the size m×n:

$$(F \otimes G)_{i,j} = \mathrm{F_{i,j}} \times \mathrm{G_{i,j}} \ , \mathrm{i} = 1,2, \dots, \mathrm{m} \text{ and } \mathrm{j} = 1,2, \dots, \mathrm{n}$$

Definition 2. Denote $\oplus$ is bit-wise XOR operator on two nonnegative integers

Example: $5 \oplus 12 = 0101 \oplus 1100 = 1001 = 9$

Definition 3. For every nonnegative integer matrix D, XSUM(D) or $\sum_{i,j}^{\oplus} D_{i,j}$ is the sum by operator $\oplus$ over all component of D.

Remark 1. If $\mathrm{D}_{i,j} \in \{0,1, \dots, 2^r - 1\} \ \forall (i,j)$, then

$$XSUM(D) \in \{0,1, \dots, 2^r - 1\}$$

III. HIDING DATA ON ONE BLOCK

This section presents algorithms for hiding data on a binary matrix (block of pixels) F of size m×n by modifying one bit at most in F.

A. *Algorithm for hiding one bit*

Wu-Lee scheme [7] is known as a simple scheme for hiding data on binary images. This scheme uses a binary random matrix K of size m×n as secret key and can hide a bit b on F by modifying one bit at most of F to get a binary matrix G to satisfy the condition:

$$SUM(G \otimes K) \bmod 2 = b$$

However, this scheme can not extend to hide a string of bits. Now, we consider a new algorithm by using operator $XSUM(G \otimes K)$ instead of $SUM(G \otimes K)$ in the Wu-Lee algorithm. This algorithm could expand to hide a string of r bits.

http://sites.google.com/site/ijcsis/
ISSN 1947-5500

Algorithm 1.

This algorithm will modify at most one element of F to get a matrix G satisfying the condition:

$$XSUM(G \otimes K) = b$$

Algorithm is performed as follows:

Step 1:

- Compute $s = XSUM(F \otimes K)$
- If s=b then set G=F and stop

Otherwise go to Step 2

Step 2:

- Compute $d = s \oplus b$
- Find an element (u,v) such that $K_{u,v} = d$
- Reverse $F_{u,v}$: $F_{u,v} = 1 - F_{u,v}$
- Set G = F and stop

Remark 2. The value of d is always equal to 1, so to Step 2 are carried out, the matrix K must satisfy the condition:

$$\{1\} \subset \{K_{i,j} | i = 1, \dots, m \text{ and } j = 1, \dots, n\}$$

B. Algorithm for hiding a bit string

In this section we expand the Algorithm 1 for hiding r bits $b = b_1 b_2 \dots b_r$ in an image block F by using the matrix P for which elements are strings of r bits. In other words, the elements $P_{i,j}$ have a value from 0 to 2^r-1.

Similar to the Algorithm 1, following algorithm will change at most one element of the matrix F to obtain matrix G to satisfy the condition:

$$XSUM(G \otimes P) = b \quad (3.1)$$

Algorithm 2.

Step 1:

- Compute $s = XSUM(F \otimes P)$ (3.2)
- If s = b, set G = F and stop

Otherwise go to Step 2

Step 2:

- Compute $d = s \oplus b$
- Find an element (u,v) such that $P_{u,v} = d$
- Reverse $F_{u,v}$: $F_{u,v} = 1 - F_{u,v}$
- Set G = F and stop

Remark 3. In the above algorithm, the value of d is an integer number from 1 to 2^r -1, so to Step 2 are carried out, the matrix P must satisfy the condition:

$$\{1, \dots, 2^r - 1\} \subseteq \{P_{i,j} | i = 1, \dots, m \text{ and } j = 1, \dots, n\} \quad (3.3)$$

From the condition (3.3) it follows that

$$r \leq \lfloor \log_2(m \times n + 1) \rfloor$$

C. Example

To illustrate the contents of Algorithm 2, we consider an example for which $b=b_1b_2$ and matrices F, P are defined as follows:

$b=b_1b_2$ =10

F

1	0	0
0	1	1
0	1	1

P

10	01	00
11	01	10
11	11	01

Step 1:

- $s = XSUM(F \otimes P) = 10 \oplus 01 \oplus 10 \oplus 11 \oplus 01 = 11$
- Since $s \neq b$, go to Step 2.

Step 2:

- $d = s \oplus b = 11 \oplus 10 = 01$
- Find (u,v) for which $P_{u,v} = d = 01$. In this case, we have three choices: (1,2), (2,2) and (2,3). Choose (u,v)=(1,2)
- Reverse $F_{1,2}$: $F_{1,2}$=1-0 = 1, and set G = F.

So after hiding two bits 10 on F, we obtain G as follows:

G

1	1	0
0	1	1
0	1	1

D. Correctness of the data hiding scheme

We need to prove matrix G obtained from Algorithm 2 satisfies condition (3.1): $XSUM(G \otimes P) = b$. This is obviously true if the algorithm ends in Step 1, so we only consider the case of the algorithm ends at step 2. Then we have:

$$P_{u,v} = d = s \oplus b \in \{1, 2, \dots, 2^r - 1\} \quad (3.4)$$

$$G_{i,j} = \begin{cases} F_{i,j}, & if\ (i,j) \neq (u,v) \\ 1 - F_{i,j}, & if\ (i,j) = (u,v) \end{cases} \quad (3.5)$$

Now if set

$$s' = XSUM(G \otimes P) = \sum_{i,j}^{\oplus} G_{i,j} \times P_{i,j}$$

Then from (3.2), (3.5) and from the fact that $a \oplus a = 0$, we obtain

$$s' = \sum_{(i,j) \neq (u,v)}^{\oplus} F_{i,j} \times P_{i,j} \oplus [(1 - F_{u,v}) \times P_{u,v}]$$

http://sites.google.com/site/ijcsis/
ISSN 1947-5500

$$s' = s \oplus [F_{u,v} \times P_{u,v}] \oplus [(1 - F_{u,v}) \times P_{u,v}]$$

Since $F_{u,v} \in \{0,1\}$, it follows from (3.4) that

$$s' = s \oplus P_{u,v} = s \oplus s \oplus b = b$$

Thus we obtain condition (3.1) and correctness of the data hiding scheme is proven.

E. Algorithm 3

To improve the safety level of the Algorithm 2, we can use an integer number $q \in \{0,1,2, \dots, 2^r - 1\}$ as a second key. We calculate Algorithm 3 with content similar to the Algorithm 2 except value s is calculated by the formula:

$$s = XSUM(F \otimes P) \oplus q$$

Additionally, to restore the bit string b, instead of the formula (3.1) we will use the following formula:

$$XSUM(G \otimes P) \oplus q = b$$

We notice that matrix G in Algorithm 3 is determined from F, P, q and b. Therefore, we can see that this algorithm as a transformation T from (F, P, q, b) to G:

G = T(F,P,q,b)

IV. DATA HIDING SCHEME IN BINARY IMAGE

A. The Inputs for scheme

Below we present use of the Algorithm 3 to hide a data bit string d in a cover binary image I. To do this, we need to use a positive integer r, a matrix P of size m×n and a sequence Q of m×n integers, which satisfy the following conditions:

- $r \le \lfloor \log_2(m \times n + 1) \rfloor$
- $P_{i,j} \in \{0,1, \dots, 2^r - 1\} \forall (i,j)$
- $\{1,2, \dots, 2^r\text{-}1\} \subset \{P_{i,j} \mid i = 1, \dots, m \text{ and } j = 1, \dots, n\}$
- $Q = \{q_1, q_2, \dots, q_{m \times n}\}$ with $0 \le q_i \le 2^r - 1$

B. Algorithm for hiding data

Step 1 (Partition): Divide the binary image I into N blocks F^i of size m×n and divide the data string d into N sub-strings b^i of size r bits.

Step 2 (Hiding data in each block):

For i=1 to N do

$$\alpha = (i - 1)\ mod\ (m \times n) + 1$$

$G^i = T(F^i, P, q_\alpha, b^i)$

End for

After executing the algorithm, we get the binary image J including N blocks G^i of size m×n.

C. Algorithm for restoring data

To restore hidden data from the stego-image J (image contains hidden information) we need to know r, m, n and secret keys P, Q. The algorithm is implemented as follows:

Step 1 (Partition): Divide the stego - image J into N blocks G^i of size m×n.

Step 2 (Restoring data):

For i = 1 to N do

$$\alpha = (i - 1)\ mod\ (m \times n) + 1$$

$$b^i = XSUM(G^i \otimes P) \oplus q_\alpha$$

End for

After executing the algorithm, we obtain data string d including N sub-strings b^i of size r bits.

D. Security Analysis of the Proposed Scheme

Each data hiding scheme often uses matrices and/or number sequences as a secret key to protect the hidden data. The greater the number of key combinations, the more difficult it is for hackers to detect the secret key used. Therefore the scheme is of higher security.

The TCP scheme uses a binary m×n matrix K and a weight m×n matrix W as the secret keys. The number of combinations for K is $2^{m \times n}$ and for W is:

$$C_{2^r-1}^{m \times n} \times (2^r - 1)! \times (2^r - 1)^{m \times n - (2^r - 1)}$$

So the number of key combinations (K, W) is:

$$\pi_{KW} = 2^{m \times n} \times C_{2^r-1}^{m \times n} \times (2^r - 1)! \times (2^r - 1)^{m \times n - (2^r - 1)}$$

In [1], the authors use a binary m×n matrix K and a serial number m×n matrix O as the secret keys. Moreover, the authors pointed out that the number of combinations for O is:

$$C_{2^r-1}^{m \times n} \times (2^r - 1)! \times (2^r)^{m \times n - (2^r - 1)}$$

So the number of key combinations (K, O) is:

$$\pi_{KO} = 2^{m \times n} \times C_{2^r-1}^{m \times n} \times (2^r - 1)! \times (2^r)^{m \times n - (2^r - 1)}$$

In the proposed scheme we use an integer m×n matrix P and a sequence Q of m×n integer numbers as the secret keys. From the definition of P and Q in subsection IV.A, it follows that the number of combinations for P is:

$$C_{2^r-1}^{m \times n} \times (2^r - 1)! \times (2^r)^{m \times n - (2^r - 1)}$$

http://sites.google.com/site/ijcsis/
ISSN 1947-5500

and for Q is $2^{r\times m\times n}$, so the number of key combinations (P, Q) is:

$$\pi_{PQ} = 2^{r\times m\times n} \times C_{2^r-1}^{m\times n} \times (2^r-1)! \times (2^r)^{m\times n-(2^r-1)}$$

In applications often choose r ≥ 2, so we have:

$$\frac{\pi_{PQ}}{\pi_{KW}} > 2^{(r-1)\times m\times n} \geq 2^{m\times n}$$

$$\frac{\pi_{PQ}}{\pi_{KO}} = 2^{(r-1)\times m\times n} \geq 2^{m\times n}$$

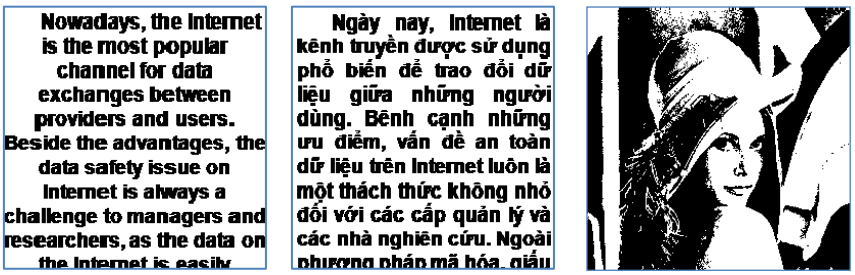
The above analysis shows that the new proposed scheme is more secure than both schemes TCP and CTL

V. EXPERIMENTS

In these experiments we use three different images of the same size 256×256 as cover images (Figure 1), including English text image, Vietnamese text image and the "Lena" image, to hide the same message with 256 bytes length (Figure 2). The data hiding in each image were performed according to two plans of dividing blocks: (m,n,r) = (8,8,6) and (m,n,r)= (16,16,8).

Table 1 presents the PSNR values of all stego-images obtained by the new scheme, the CTL scheme and the TCP scheme, respectively. The results indicate that, PSNR values of the new scheme are always higher than those of TCP scheme and the same as those of CTL scheme.

Table 2 presents number of pixels modified in each image after performing data hiding by above schemes. The results indicate that these numbers of the new scheme are always smaller than those of TCP scheme and the same as those of CTL scheme.



(a) (b) (c)

Fig. 1. Cover images: (a) English text image, (b) Vietnamese text image, (c) Lena image

It is important to understand that cyber warfare does not necessarily have anything to do with the internet. Many of the more devastating cyber - attacks can not be launched remotely, as the most critical networks are not connected to the public network.

Fig. 2. The secret message with 256 characters

Table 1. PSNR values of stego-images of three schemes

Block size / Cover Image	8×8			16×16		
	New scheme	*CTL scheme*	*TCP scheme*	*New scheme*	*CTL scheme*	*TCP scheme*
Vietnamese text image	22,901 dB	**22,94 dB**	21,83 dB	24,116 dB	**24,134 dB**	23,196 dB
English text image	**22,94 dB**	**22,94 dB**	22,005 dB	**24,134 dB**	24,116 dB	23,1 dB
Lena image	**22,901 dB**	22,889 dB	22,166 dB	**24,151 dB**	**24,151 dB**	22,967 dB

Table 2. Number of modified pixels in stego images of three schemes

Block size / Stego Images	8×8			16×16		
	New scheme	*CTL scheme*	*TCP scheme*	*New scheme*	*CTL scheme*	*TCP scheme*
Vietnamese text image	336 bits	**333 bits**	430 bits	254 bits	**253 bits**	314 bits
English text image	**333 bits**	**333 bits**	413 bits	**253 bits**	254 bits	321 bits
Lena image	**336 bits**	337 bits	398 bits	**252 bits**	**252 bits**	331 bits

http://sites.google.com/site/ijcsis/
ISSN 1947-5500

VI. CONCLUSIONS

This paper presents a new scheme for embedding secret data into a binary image. For each block of m × n pixels, the new scheme can hide $\lfloor log_2(m \times n + 1) \rfloor$ bits of data by changing one bit at most in block. The experimental results indicate that if embedding a same amount of secret data in a same cover image, the stego-image quality of the new scheme is similar to that of CTL scheme and better than that of TCP scheme. The theoretical analyses have confirmed that the new proposed scheme is indeed more secure than both schemes TCP and CTL. Additionally, as compared to two schemes above, the new scheme is simpler and easier to install for applications.

REFERNCES

[1] Chin-Chen Chang, Chun-Sen Tseng, Chia-Chen Lin. *"Hiding Data in Binary Images"*, ISPEC 2005, LNCS 3439, pp 338-349, 2005.

[2] Guo Fu Gui, Ling Ge Jiang, and Chen He, *"A New Asymmetric Watermarking Scheme for Copyright Protection",* IECE Trans. Fundamentals, Vol. E89-A, No. 2 February 2006.

[3] Y. K. Lee and L. H. Chen, "High Capacity Image Steganographic Model," in *Proc. of IEE International Conference on Vision, Image and Signal Processing*, Vol. 147, No. 3, pp.288-294 (2000).

[4] B. Smitha and K.A. Navas, "Spatial Domain – High Capacity Data Hiding in ROI Images", IEEE – ICSCN 2007, MIT Campus, Anna University, Chennai, India, Feb, 22-24,2007. pp.528-533.

[5] Y.C. Tseng, Y. Y. Chen, and K. H. Pan, *"A secure Data Hiding Scheme for Binary Images"*, IEEE Transactions on Communications, Vol. 50, No. 8, August, pp. 1227-1231 (2002) Symposium On Computer and Communication, 2000.

[6] C. H. Tzeng, Z. F. Yang, and W. H. Tsai. "Adaptive Data Hiding in Palette Image by Color Ordering and Mapping with Security Protection," *IEEE Transactions on Communications*, Vol. 52, No. 5, May, pp. 791- 800 (2004)

[7] M. Y. Wu and J. H. Lee, "A Novel Data Embedding Method for Two-color Facsimile Images," in *Proc. Int. Symp. on Multimedia Information Processing*, Chung-Li, Taiwan, R.O.C., Dec. (1998).

[8] J. Zhao and E. Koch, "Embedding Robust Labels into Images for Copyright Protection," in *Proc. Int. Conf. Intellectual Property Rights for Information Knowledge*, New Techniques,

AUTHORS PROFILE

Do Van Tuan received M.Sc. degree in Information Technology in 2007 from Vietnam National University, Ha Noi. He is currently a PhD student at Hanoi University of Science and Technology. His research interests include Data Hiding, Digital Watermarking, Cryptography

Tran Dang Hien received M.Sc. degree in Information Technology in 2010 from Vietnam National University, Ha Noi. He is currently a PhD student at Vietnam National University. His research interests include Data Hiding, Digital Watermarking, Image Forensic.

Pham Van At received B.Sc. and PhD degree in Mathematics in 1967 and 1980 from Vietnam National University, Ha Noi. Since 1984 he is Associate Professor at Faculty of Information Technology of Hanoi University of Transport and Communication. His research interests include Linear algebra, optimization, Image processing, Data Hiding, Cryptography.

http://sites.google.com/site/ijcsis/
ISSN 1947-5500

Virtual Investigation of Patients for Medical E-Learning

A. M. Riad

Faculty of Computer and Information, Sciences, Mansoura University
Mansoura, Egypt

Hazem M. El-Bakry

Faculty of Computer and Information, Sciences, Mansoura University
Mansoura, Egypt
helbakry20@yahoo.com

Samir M. Abd El-razek

Faculty of Computer and Information, Sciences, Mansoura University
Mansoura, Egypt

Abstract— The main objective of this research is to allow medicine students to practice problem-based learning by simulating a medical consultation, using the current pedagogical methods based on clinical cases and integrating web technology methods. In a previous paper [11], a virtual medical e-learning system for a few number of students. In this paper the design is introduced to combine large number of students. The design is applied for Egyptian Universities Virtual Environment System (EUVES), a virtual environment for medical e-learning for all medicine students in the Egyptian universities. The system is logically designed to simulate case-based learning methodology, and technically designed upon MedBiquitous specifications for virtual patient. In this paper the current researches of virtual patients and MedBiquitous specifications which became a major standard for building virtual patients are introduced. Then the concept of case-based learning and the need for virtual environment is presented to support this methodology of learning. Finally an overview of workflow, data structure, and functional modules of EUVES are given.

Keywords- Virtual Environment, Virtual Patient, Case Based Learning, Medical E-Learning.

I. INTRODUCTION

Recent studies of medical learning methodologies have a major interest for case-based learning (CBL). This methodology of learning depends on putting students in a semi-real situation and the objective is to learn how to react in this particular situation [1]. Medical cases are based on real scenarios and supporting data to evaluate an open ended problem. Finding a real patient to demonstrate all medical case would be a difficult job for professors. Here comes the value of virtualization which enables the teacher to create a virtual case based on a predefined problem and known answer. Many researchers developed frameworks for virtualizing medical cases using virtual reality technologies, which gained some success factors for simulating patients, surgeries, and clinical tools. Virtual reality systems have some rich graphical features and 3D engines containing great numbers of graphical elements and complex processing and rendering which are not suitable for the current trends for using the web technology in e-learning. In the last few years new web-based models and frameworks have been developed to make use of the accessibility, scalability, and low cost for using web. In this paper we introduce Egyptian Universities Virtual Environment System (EUVES), a new web based virtual environment system for medical CBL based on Virtual Patients (VPs). The use of virtual patient is increasing in healthcare education, partly in response to increasing demands on health care professionals and education of students but also because they allow opportunity for students to practice in a safe environment.

II. VIRTUAL PATIENT: CURRENT RESEARCHES

The European Commission is funding a major 3-years project named eViP [2], the project started in 2007 and is due to finish by end of 2010. The project aims to explore the feasibility of repurposing and enriching VP examples in a variety of different ways, and for different purposes [3]. The project is developed with cooperation with seven universities and they could implement common technical standards for all VPs in collaboration with MedBiquitous. MedBiquitous Virtual Patients Working Group has developed XML standards and Web services requirements to enable interoperability, accessibility and reusability of Web-based virtual patient learning content [4]. In 2009 Inga Hege from the University of Muenchen-Germany has implemented the MedBiquitous standards into a learning system named CASUS [5]. Another project has been developed in the University of Maryland-USA, Marjorie and Nirenburg has been working on a simulation and a training system the name it Maryland Virtual Patient (MVP). The system gives the user the ability to act as a physician and treats virtual patients with or without a virtual totur [6]. One of the major projects that are working on VPs is CAMPUS. The CAMPUS virtual patient system, developed at the University of Heidelberg-Germany consists of different modules for learning and assessment with VPs [7]. Some

http://sites.google.com/site/ijcsis/
ISSN 1947-5500

projects has a big focus of authoring VPs like VPSIM that has been developed in the University of Pittsburg-USA on 2008. VPSIM creates an authoring environment for medical educators, that requires minimal training and stimulated creative case writing [8]. Also researchers in the University of McGill-Canada have developed an authoring system to create VPs can be used in the simulation of physical examination, taking history, and ordering the required laboratory tests [9].

III. MEDBIQUITOUS VIRTUAL PATIENT (MVP) SPECIFICATIONS

The MedBiquitous Consortium developed the XML-based MedBiquitous Virtual Patient (MVP) data specification. This specification enables the exchange and reuse of VPs across different systems. The following are the main components of the MVP architecture [**Error! Bookmark not defined.**], shown in fig 1:

The Virtual Patient Data (VPD) provides the personal and clinical data that is related to the clinical scenario being simulated. The VPD contains data elements and some kind of structure that represents the medical history, physical and technical examination and therapy. The Data Availability Model (DAM) specifies the relation between the VPD and Media Resources (MR) elements to be used in the Activity Model. VPD and MR elements can be used multiple times in the DAM nodes depending on the logical relation between these elements, such as patient history and test results. The Activity Model (AM) encodes what the learner can do and how he may interact with the virtual patient. By creating available paths through content using interconnected nodes and controlling how the user can follow them using a simple rule system, a great variety of patient activities are possible. The AM provides the context in which they are exposed to the learner. Media Resources (MR) like images, animations, videos and audio files associated with the virtual patient are referenced in the DAM or in the VPD elements. IMS Content Packaging is used to structure media resources within the MVP specification and provide unique identifier for each media resource [10].

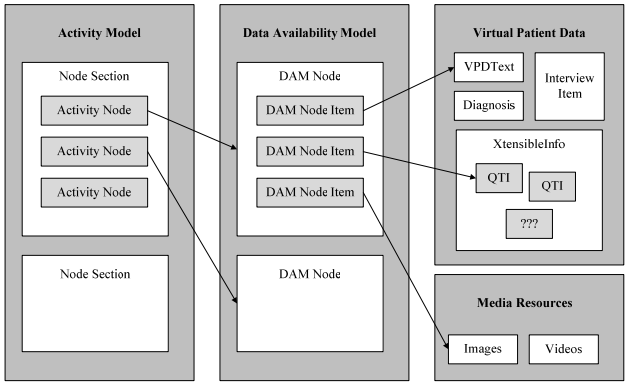


Fig. 1. MVP Specification Model

IV. OVERVIEW OF EUVES

EUVES is a medical e-learning system for authoring and delivering CBL to medicine students based on MVP standards. EUVES can work as helper tool for medicine professors to demonstrate the diagnosis process for students over the web in a virtual lecture style presented in fig 2. Professors can select or create a new VP template, these templates are categorized by teacher upon their relevance to subjects. Each VP template contains a complete VP data and media resources, the teacher can create an instance from this template to start performing the diagnosis process on this instance. The idea of creating templates is to "reuse" previously created VPs and make it easier for creating VPs rather than copying files. After creating the instance the teacher may need to change some VP data for more clarification on this case or to add new important information needed for diagnosis like up-normal values for some lab elements, or increase the measure of blood pressure, etc.

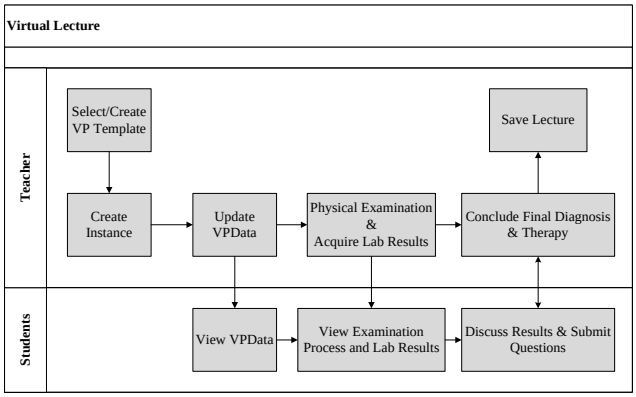


Fig. 2. Virtual Lecture Workflow [11]

All the previous steps are performed "offline" or away of students. Once the teacher finishes his updates on the instance chosen, he can start the virtual lecture. The virtual lecture items are broadcasted to students over the web, starting with VP data and media resources. Teacher starts his lecture using collaboration tools like audio/video streaming, whiteboard, and performing visible physical examination on the VP instance. The physical examination is performed using a rich set of physician tools, and possibly to acquire some lab results for this VP instance. Each step of the diagnosis process can be opened for discussion by students using a simple chatting tool with the teacher until the teacher reach the conclusion of the diagnosis student can submit their questions and discuss the reasonability of diagnosis. The lecture can be saved to the Virtual Lecturers Library (VLL) and can be reviewed by student later. Another important feature of EUVES is it can be used as an assessment tool for students. Teacher can "setup" the instance as introduced previously, but he will assign the diagnosis process to a group of students who need to collaborate together to conclude the possible diagnosis with guidance of the teacher.

http://sites.google.com/site/ijcsis/
ISSN 1947-5500

V. EUVES DATA STRUCTURE

In EUVES, VP's XML data structure is based MedBiquitous VP specifications. This global data structure enables EUVES to integrate with any other VP authoring system by importing VP XML files. An XML package is a set of XML files that represents a single VP. This standardization required implementing two components for building VP XML packages, and another for validating XML files schema before loading in the virtual lecture player as shown in fig. 3 [11].

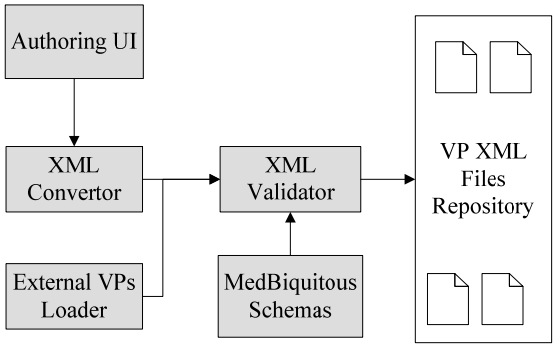


Fig. 3. VPs XML Handlers

VP data structure is defined by a root element named VirtualPatientData, which contains the clinical and demographic data for VPs in addition to the metadata. VirtualPatientData element must exist once in a VP package. VirtualPatientData has sub elements defining the data of a VP shown in fig 4, the existence of any sub element is optional [10].

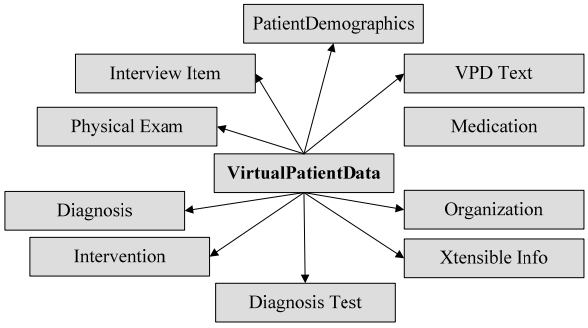


Fig. 4. VP data structure elements

A. Patient Demographics

The PatientDemographics element contains subelements that define the name, age, sex, and other demographic characteristics of a virtual patient and allows for the categorization or grouping of various demographic characteristics, see the example shown in fig. 5. It consists of two sub elements: Core Demographics which contains the regular demographics, and Demographics Characteristic which contains non-standard demographics.

```
<PatientDemographics id="D009">
    <CoreDemographics id="C1003">
        <PatientID> P542 </PatientID>
        <Name> Adam Ali </Name>
        <Age> 50 </Age>
        <Gender> Male </Gender>
    </CoreDemographics>
    <DemographicCharacteristic id="154B">
        <Title> Marital Status </Title>
        <Description> Single </Description>
    </DemographicCharacteristic>
</PatientDemographics>
```

Fig. 5. Patient Demographics Example

B. VPD Text

It provides narrative or other descriptive text that is part of the virtual patient data. It contains A unique identifier for this text that can be used by other virtual patient components to control how the text is used in an educational activity, also it contains an indicator of the function this text serves in the virtual patient. Valid values are: complaint, history, or problem. VPD Text could contain an xhtml div element that can be used to format how text will be presented to the user.

C. Medication

Medication describes a medication taken by the virtual patient in detail, including the medication name, dose, route, and frequency.
Medication name defines the name of the medication, possibly referencing a medical taxonomy or vocabulary. Dose defines the dosage of the medication, for example: 10 mg. Route defines the route of administration for the medication, for example: oral.

D. Interview Item

It describe a single question and response that is part of a clinical history, see the example shown in fig 6.

```
<InterviewItem id="1">
    <Question>How are you feeling?</Question>
    <Response>Not very well.</Response>
</InterviewItem>
```

Fig. 6. Interview Item Example

E. Physical Exam

It describes a single physical exam and the findings of that exam. It consists of sub elements like exam name, location on body, actions the virtual clinician takes to perform the exam,

http://sites.google.com/site/ijcsis/
ISSN 1947-5500

findings of the exam for example: normal heartbeat, and the description of the physical exam in more details.
Location on body is defined by sub elements like which part of body, whether it's proximal or distal, right or left, front or back, inferior or superior.

F. Diagnostic Test

Diagnostic test defines test result, it consists of test name, unit to measure this test, result, and the normal value of the test.

G. Diagnosis

Diagnosis defines either a single differential diagnosis or a single final diagnosis intended by the virtual patient author. It contains the likelihood of the differential diagnosis being correct. Valid values are: high, medium, low, none.

VI. EUVES MODULES

EUVES is composed of main four functional modules: configuration module, learning management module, VP authoring module, and virtual lecture module. EUVES supports three types of users, with different authorization levels: student, teacher, and administrator.

A. Configuration module

The administrator module manages users, permissions, and lookups. System administrator is responsible for managing users of the system and defining their permissions upon their types.

B. Learning management module

Professors can access to VPs templates databases and they can create user specific virtual cases. Each teacher has his own set of virtual cases and then he can use these cases in virtual lectures or assessments. Professors also can create new assessments and select students who will perform this virtual assessment. Through this module professors can categorize virtual lectures into subjects and can view reports related to each subject like number of virtual lectures, or assessments reports.

C. VP authoring Module

This module provides a tool for building VPs with a simple user interface for professors who does not need to know about the structure of VP, without help of multimedia developers or designers. Most of VPs created on EUVES are built using VP templates which were developed with assistance of Mansoura Medicine Faculty professors who provided the initial information needed for each case. Based on the MVP specification model, templates contain VP data like text, media resources, diagnosis, some questions and answers. This information must lead to a specific medical conclusion about this VP. All templates are based on the normal template that presents the normal conditions and data for a normal human. Then professors use these templates to define their own virtual cases by updating some information of the template that could makes it easier for students to perform the diagnosis or by making it harder in assessments. Professors can update VP data from a usable web interface shown in fig. 7. VP data in authoring module are categorized into general information, patient history, physical examination, lab tests, diagnosis, and therapy.

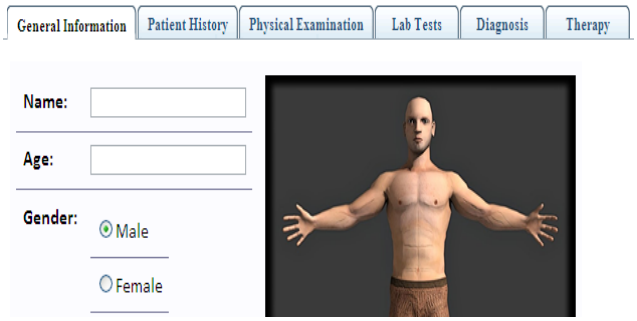


Fig. 7. VP authoring Module

Professors use these categories for submitting VP data, which consists mainly of text, boolean, images, and different multimedia formats. History information, physical examination, and lab tests are considered important to solve the case can be marked as required. The amount of time needed to create a single case varies and may depend on the complexity of the case.

D. Virtual lecture module

This module is main core of EUVES. It manages the CBL concept of putting users in a virtual situation, and starts to record and analyze their activities for solving this situation. The main components of this module are shown in fig. 8.

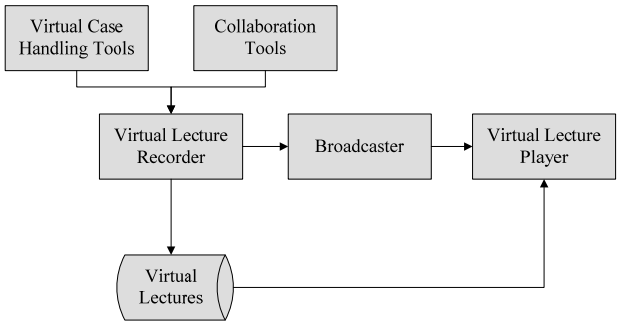


Fig. 8. Virtual lecture module components

The virtual case handling tools are an important component of this module, it manages the VP data and the virtual activities

http://sites.google.com/site/ijcsis/
ISSN 1947-5500

like diagnosis and lab tests. An example of the virtual activities is shown in fig. 9.

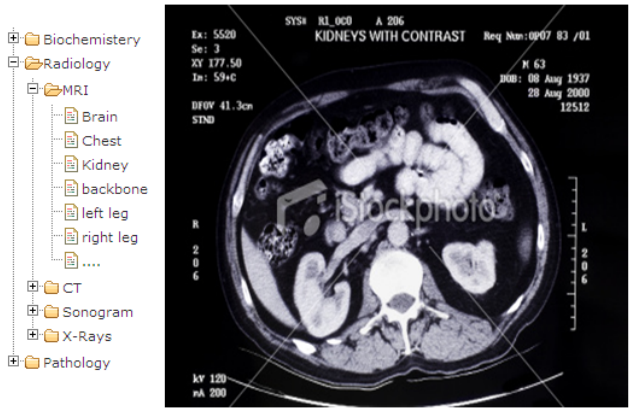


Fig. 9 Virtual Activities

Some VP data are visible to users of EUVES, others are hidden waiting for virtual activities to be performed to reveal these hidden information about the virtual case. In e-learning context collaboration tools are very important to help professors and students to share knowledge and for open discussions related to virtual cases like discussion boards, whiteboard, and chatting.

Both virtual activities and collaboration activities are recorded using the virtual lecture recorder as a one unit like a video recording for a real lecture. Students can view any recorded virtual lectures from the virtual lectures database. Live virtual lectures are broadcasted to students and viewed using the virtual lecture player, which is a web-based interface that presents virtual case properties, virtual activities, and collaboration activities.

EUVES provides another feature which is virtual assessment. In virtual assessments students plays the physician role, and they conclude the final results and therapy plan. Then teacher evaluate these results like the real assessment scenario. Virtual assessment is built on the same components of this module.

VII. Conclusion

EUVES consists of different modules for learning and assessment with virtual patients based on case-based learning methodology. An easy to use authoring system enables the user to create virtual patients based on a defined structure of virtual patient data like patient demographics, interviews, diagnosis, and medications. Building the system on MedBiquitous specifications enables EUVES to integrate with any other system built on the same specifications like eVip by importing and exporting virtual patients XML files. Medicine professors can use EUVES to practice virtual diagnosis and therapy in a virtual lecture, or assess students online. Students can communicate with the lecturer by a rich collaboration tools like whiteboard, and discussion boards. Virtual lectures can be recorded containing virtual case, virtual activities, students' questions and answers.

References

[1] Samawal Hakim, "Learning methods and its application in medical education", 2007
[2] http://www.virtualpatients.eu/about/about-evip , Last Accessed on July 20, 2010
[3] eViP Project Team, "Annual Report" , 2009
[4] http://www.medbiq.org , Last Accessed on July 29, 2010
[5] Inga Hege, A. Kononowicz, M. Pfahler, "Implementation of the MedBiquitous Standards into The Learning System CASUS" , 2009
[6] Marjorie McShane, Sergei Nirenburg, "Maryland Virtual Patient: A Knowledge-Based Language-Enabled Simulation and Training System", 2009
[7] Jorn Heid, Benjamin Hanbeck, Frank Hess, "Demonstration of the CAMPUS Virtual Patient System" , 2009
[8] James Mcgee, "VPSIM – A Standards-Based Virtual Patient Authoring System", 2009
[9] David Fleiszer, Nancy Posel, "Open Virtual Patient Authoring System From McGill University", 2010.
[10] Valerie Smothers, Ben Azan, Rachel Ellaway, "MedBiquitous Virtual Patient Specifications and Description Document", 2010
[11] Samir M. Abd El-Razek, Waeil F. Abd El-Wahed and, Hazem M. El-Bakry, "EUVES: A Virtual Environment System for Medical Case Based Learning," International Journal of Computer Science and Network Security, vol. 10, no. 9, September 2010, pp. 159-163.

http://sites.google.com/site/ijcsis/
ISSN 1947-5500

Effect of AWGN & Fading (Raleigh & Rician) channels on BER performance of a WiMAX communication System

Nuzhat Tasneem Awon
Dept. of Information & Communication Engineering
University of Rajshahi, Rajshahi, Bangladesh
e-mail: tasneemawon@gmail.com

Md. Mizanur Rahman
Dept. of Information & Communication Engineering
University of Rajshahi, Rajshahi, Bangladesh
e-mail: mizan5624@yahoo.com

Md. Ashraful Islam
Lecturer
Dept. of Information & Communication Engineering
University of Rajshahi, Rajshahi, Bangladesh
e-mail: ras5615@gmail.com

A.Z.M. Touhidul Islam
Associate Professor
Dept. of Information & Communication Engineering
University of Rajshahi, Rajshahi, Bangladesh
e-mail: touhid_ict_it@yahoo.com

Abstract— **The emergence of WIMAX has attracted significant interests from all fields of wireless communications including students, researchers, system engineers and operators. The WIMAX can also be considered to be the main technology in the implementation of other networks like wireless sensor networks. Developing an understanding of the WIMAX system can be achieved by looking at the model of the WIMAX system. This paper discusses the model building of the WIMAX physical layer using computer MATLAB 7.5 versions. This model is a useful tool for BER (Bit error rate) performance evaluation for the real data communication by the WIMAX physical layer under different communication channels AWGN and fading channel (Rayleigh and Rician), different channel encoding rates and digital modulation schemes which is described in this paper. This paper investigates the effect of communication channels of IEEE 802.16 OFDM based WIMAX Physical Layer. The performance measures we presented in this paper are: the bit error rate (BER) versus the ratio of bit energy to noise power spectral density (Eb/No). The system parameters used in this paper are based on IEEE 802.16 standards. The simulation model built for this research work, demonstrates that AWGN channel has better performance than Rayleigh and Rician fading channels. Synthetic data is used to simulate this research work.**

Keywords-WiMAX;Communication Channel;CRC Codind; styling; insert (key words)

I. INTRODUCTION

The wireless broadband technologies are bringing the broadband experience closes to a wireless context to their subscribers by providing certain features, convenience and unique benefits. These broadband services can be categorized into two types; Fixed Wireless Broadband and Mobile Broadband. The fixed wireless broadband provides services that are similar to the services offered by the fixed line broadband. But wireless medium is used for fixed wireless broadband and that is their only difference. The mobile broadband offers broadband services with an addition namely the concept of mobility and nomadicity. The term nomadicity can be defined as "Ability to establish the connection with the network from different locations via different base stations" while mobility is "the ability to keep ongoing connections engaged and active while moving at vehicular speeds". Examples of wireless broadband technologies are Wireless LAN and WIMAX.

WIMAX is the abbreviation of Worldwide Interoperability for Microwave Access and is based on Wireless Metropolitan Area Networking (WMAN). The WMAN standard has been developed by the IEEE 802.16 group which is also adopted by European Telecommunication Standard Institute (ETSI) in High Performance Radio Metropolitan Area Network, i.e., the HiperMAN group. The main purpose of WIMAX is to provide broadband facilities by using wireless communication [1]. WIMAX is also known as "Last Mile" broadband wireless access technology WIMAX gives an alternate and better solution compared to cable, DSL and Wi-Fi technologies as depicted in Figure-a: [2]

http://sites.google.com/site/ijcsis/
ISSN 1947-5500

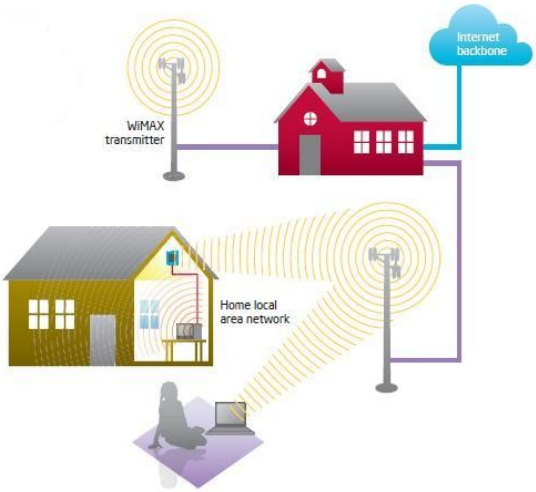


Figure-a: WiMAX System

Like other wireless communication network, transmission medium faces two major problems in WIMAX communication system. These problems are:
a) AWGN noise &
b) Rayleigh and Rician Fading.

AWGN noise
AWGN is a noise that affects the transmitted signal when it passes through the channel. It contains a uniform continuous frequency spectrum over a particular frequency band.

Rayleigh Fading
When no LOS path exists in between transmitter and receiver, but only have indirect path than the resultant signal received at the receiver will be the sum of all the reflected and scattered waves.

Rician Fading
It occurs when there is a LOS as well as the non-LOS path in between the transmitter and receiver, i.e. the received signal comprises on both the direct and scattered multipath waves. [2]

The objective of this project is to implement and simulate the IEEE 802.16 OFDM based WiMAX Physical Layer using MATLAB in order to have a better understanding of the standards and evaluate the system performance based on the effect of different communication channels. This involves studying through simulation, the various PHY modulations, coding schemes and evaluating the bit error rate (BER) performance of the WIMAX communication system under different channel models such as, AWGN channel and Fading (Rayleigh & Rician) channels.

II. Simulation Model

The transmitter and receiver sections of the WiMAX Physical layer are shown in the block diagram of Figure-b. This structure corresponds to the physical layer of the WiMAX air interface. In this setup, we have just implemented the mandatory features of the specification, while leaving the implementation of optional features for future work. The channel coding part is composed of coding techniques of the Cyclic Redundancy Check (CRC) and Convolutional Code (CC). The complementary operations are applied in the reverse order at channel decoding in the receiver end. We do not explain each block in details. Here we only give the emphasis on communication channel i.e. AWGN and Fading (Rayleigh and Rician) and Cyclic Redundancy Check (CRC) and Convolutional Code (CC) coding techniques.

A Convolution encoder consists of a shift register which provides temporary storage and a shifting operation for the input bits and exclusive-OR logic circuits which generate the coded output from the bits currently held in the shift register. In general, k data bits may be shifted into the register at once, and n code bits generated. In practice, it is often the case that k=1 and n=2, giving rise to a rate 1/2 code [3].

Cyclic Redundancy Check (CRC) codes are a subset of then class of linear codes, which satisfy the cyclic shift property such as if C=[C_{n-1} ,C_{n-2} ……,C_O] is a codeword of a cyclic code, then [C_{n-2} C_{n-2} ,…,C_0,C_{n-1}], obtained by a cyclic shifts of the elements of C, is also a code word. In other word all cyclic shifts of C are code words. From the cyclic property, the codes possess a great deal of structure which is exploited to greatly simplify the encoding and decoding operation [4].

Reasonable assumption for a fixed, LOS wireless channel is the additive white Gaussian noise (AWGN) channel [5], which is flat and not "frequency-selective" as in the case of the fading channel. Particularly fast, deep frequency-selective fading as often observed in mobile communications is not considered in this thesis, since the transmitter and receiver are both fixed. This type of channel delays the signal and corrupts it with AWGN. The AWGN is assumed to have a constant PSD over the channel bandwidth, and a Gaussian amplitude probability density function. This Gaussian noise is added to the transmitted signal prior to the reception at the receiver as shown in Figure-c [6], therefore the transmitted signal, white Gaussian noise and received signal are expressed by the following equation with s(t), n(t) and r(t) representing those signals respectively:

$$r(t)=s(t)+n(t)$$

http://sites.google.com/site/ijcsis/
ISSN 1947-5500

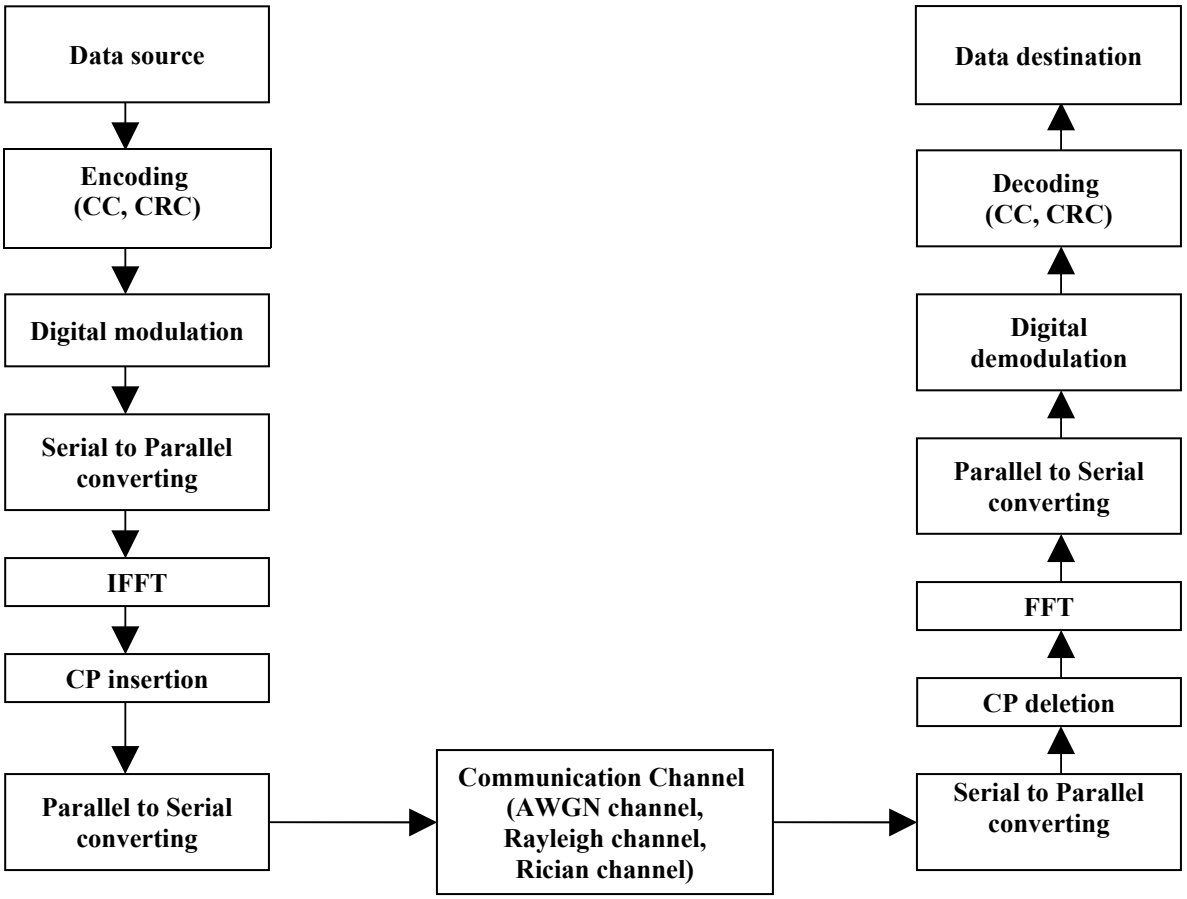


Figure-b: A block diagram for WIMAX Communication system

http://sites.google.com/site/ijcsis/
ISSN 1947-5500

Where n(t) is a sample function of the AWGN process with probability density function (pdf) and power spectral density **[7].**

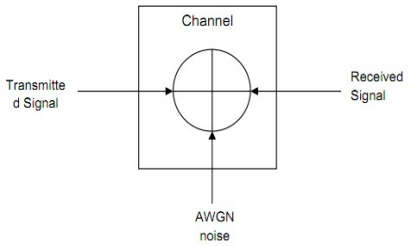


Figure-c: AWGN channel model

The in-phase and quadrature components of the AWGN are assumed to be statistically independent, stationary Gaussian noise process with zero mean and two-sided PSD of NO/2 Watts/Hz. As zero-mean Gaussian noise is completely characterized by its variance, this model is particularly simple to use in the detection of signals and in the design of optimum receivers [6]. So, it was developed using 'awgn' function which is also available in Matlab.

Multipath fading results in fluctuations of the signal amplitude because of the addition of signals arriving with different phases. This phase difference is caused due to the fact that signals have traveled different distances by traveling along different paths. Because the phases of the arriving paths are changing rapidly, the received signal amplitude undergoes rapid fluctuation that is often modeled as a random variable with a particular distribution.
The most commonly used distribution for multipath fast fading is the Rayleigh distribution, whose probability density function (pdf) is given by

$$f_{ray}(r) = \frac{r}{\sigma^2}\exp\left(-\frac{r^2}{2\sigma^2}\right), \quad r \geq 0$$

Here, it is assumed that all signals suffer nearly the same attenuation, but arrive with different phases. The random variable corresponding to the signal amplitude is r. Here σ2 is the variance of the in-phase and quadrature components. Theoretical considerations indicate that the sum of such signals will result in the amplitude having the Rayleigh distribution of the above equation . This is also supported by measurements at various frequencies. The phase of the complex envelope of the received signal is normally assumed to be uniformly distributed in [0,2π].

When strong LOS signal components also exist, the distribution is found to be Rician, the pdf of such function is given by:

$$f_{ric}(r) = \frac{r}{\sigma^2}\exp\left(\frac{-\left(r^2+A^2\right)}{2\sigma^2}\right)I_0\left(\frac{Ar}{\sigma^2}\right), \quad r \geq 0, A \geq 0$$

Where σ^2 is the variance of the in-phase and quadrature components. A is the amplitudeof the signal of the dominant path and I0 is the zero-order modified Bessel function of the first kind. Normally the dominant path significantly reduces the depth of fading, and in terms of BER Ricean fading provides superior performance to Rayleigh fading. The probability of having line-of-sight (LOS) component depends on the size of the cell. The smaller the cell the higher the probability of having LOS path. If there is no dominant path then the Rician pdf reduces to Rayleigh pdf. When A is large compared with σ, the distribution is approximately Gaussian. Thus, since Ricean distribution covers also Gaussian and Rayleigh distribution, mathematically the Ricean fading channel can be considered to be general case [8].

The procedures that we have followed to develop the WiMAX physical layer simulator is briefly stated as follows:
At the transmission section:

1. At first we have generated a random data stream of length 44000 bit as our input binary data using Matlab 7.5. Then randomization process has been carried out to scramble the data in order to convert long sequences of 0's or 1's in a random sequence to improve the coding performance.
2. Secondly we have performed Cyclic Redundancy Check (CRC) encoding. After this 1/2 rated convolutional encoding is also implemented on the CRC encoded data. The encoding section was completed by interleaving the encoded data.
3. Then various digital modulation techniques, as specified in WiMAX Physical layer namely QAM, 16-QAM and 64-QAM are used to modulate the encoded data.
4. The modulated data in the frequency domain is then converted into time domain data by performing IFFT on it.
5. For reducing inter-symbol interference (ISI) cyclic prefix has been added with the time domain data.
6. Finally the modulated parallel data were converted into serial data stream and transmitted through different communication channels.
7. Using Matlab built-in functions, "awgn", "rayleighchan" and "ricianchan" we have generated AWGN, Rayleigh and Rician channels respectively.

At the receiving section we have just reversed the procedures that we have performed at the transmission section. After ensuring that the WiMAX PHY layer simulator is working properly we started to evaluate the performance of our

http://sites.google.com/site/ijcsis/
ISSN 1947-5500

developed system. For this purpose we have varied encoding techniques and digital modulation schemes under AWGN and frequency-flat fading (Rayleigh/ Rician) channels. Bit Error Rate (BER) calculation against different Signal-to-Noise ratio (SNR) was adopted to evaluate the performance.

The simulation Parameters used in the present study are shown in Table 1.

Table 1: Simulation Parameters

Parameters	values
Number Of Bits	44000
Number Of Subscribers	200
FFT Size	256
CP	1/4
Coding	Convolutional Coding(CC), Cyclic redundancy Check (CRC)
Code rate	CC(1/2) ,CRC(2/3)
Constraint length	7
K-factor	3
Maximum Doppler shift	100/40Hz
SNR	0-30
Modulation	QAM, 16-QAM, 64-QAM
Noise Channels	AWGN, Rayleigh and Rician

III. Simulation result

This section of the chapter presents and discusses all of the results obtained by the computer simulation program written in Matlab7.5, following the analytical approach of a wireless communication system considering AWGN, Rayleigh Fading and Rician Fading channel. A test case is considered with the synthetically generated data. The results are represented in terms of bit energy to noise power spectral density ratio (Eb/No) and bit error rate (BER) for practical values of system parameters.

By varying SNR, the plot of Eb/No vs. BER was drawn with the help of "semilogy" function. The Bit Error Rate (BER) plot obtained in the performance analysis showed that model works well on Signal to Noise Ratio (SNR) less than 25 dB. Simulation results in figure 5.1, figure 5.2 and figure 5.3 shows the performance of the system over AWGN and fading (Rayleigh & Rician) channels using QAM, 16-QAM and 64-QAM modulation schemes respectively.

Performance of OFDM based WIMAX Physical layer using QAM modulation technique:

The following figure shows the BER performance of WIMAX Physical layer through AWGN channel, Rayleigh and Rician fading channels using Quadrature Amplitude Modulation (QAM) technique. The effect of AWGN channel and fading (Rayleigh & Rician) channels, we get through this figure has been discussed later.

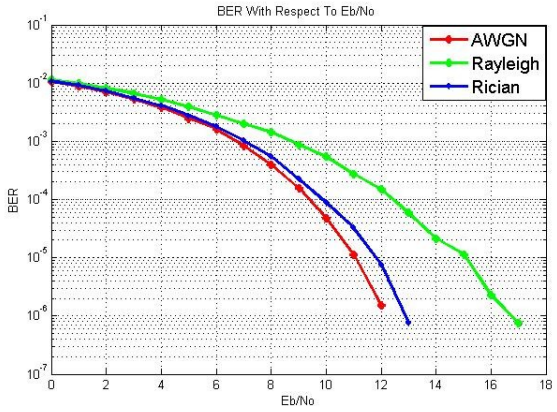


Figure-d: Bit error rate (BER) performance of AWGN, Raleigh and Rician channels for QAM modulation technique.

Performance of OFDM based WIMAX Physical layer using 16-QAM modulation technique:

The following figure shows the BER performance of WIMAX Physical layer through AWGN channel, Rayleigh and Rician fading channels using 16-QAM technique. The effect of AWGN channel and fading (Rayleigh & Rician) channels, we get through this figure has been discussed later.

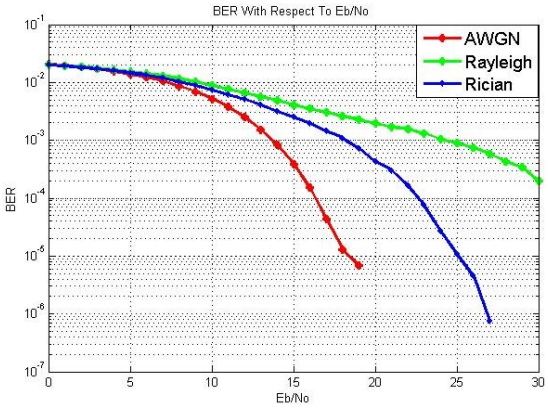


Figure-e: Bit error rate (BER) performance of AWGN, Raleigh and Rician channels for 16-QAM modulation technique.

http://sites.google.com/site/ijcsis/
ISSN 1947-5500

Performance of OFDM based WIMAX Physical layer using 64-QAM modulation technique:

The following figure shows the BER performance of WIMAX Physical layer through AWGN channel, Rayleigh and Rician fading channels using 64-QAM technique. The effect of AWGN channel and fading (Rayleigh & Rician) channels, we get through this figure has been discussed later.

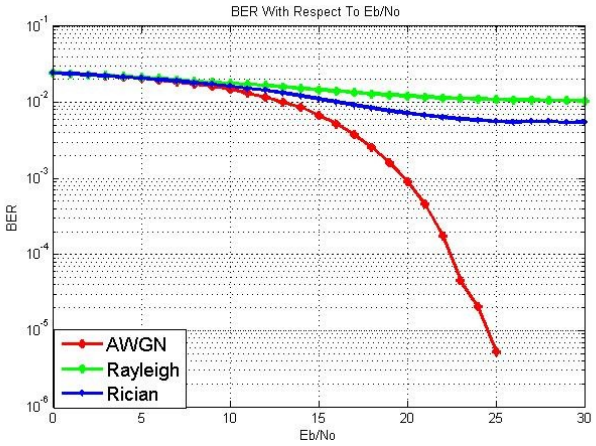


Figure-f: Bit error rate (BER) performance of AWGN, Raleigh and Rician channels for 64-QAM modulation technique.

Effect of AWGN channel on BER performance of WIMAX Physical layer:

From figure-d, e & f, we can see that, AWGN channel has lower BER than Raleigh and Rician fading channel. For an example, while using the QAM modulation scheme, for SNR value 13, BER for AWGN channel remains 0, where BER for Rayleigh and Rician channel remains 5.9091e-05 and 7.5758e-07 respectively. After SNR value 13, BER for AWGN remains zero for the rest of the SNR values. But Raleigh & Rician fading channel has more non-zero BER values than that of AWGN channel

Effect of Raleigh fading channel on BER performance of WIMAX Physical layer:

From figure d, e & f, we can see that, Raleigh fading channel has higher BER than AWGN and Rician fading channel. For an example, while using the QAM modulation scheme, for SNR value 17, BER for Raleigh fading channel remains 7.5758e-07, where BER for both AWGN and Rician channel remains zero. After SNR value 12 and after SNR value 13, BER for AWGN and for Rician fading channel remains zero for the rest of the SNR values, where Rayleigh fading channel has more non-zero BER values.

Effect of Rician fading channel on BER performance of WIMAX Physical layer:

From figure d, e & f, we can see that, Rician fading channel has higher bit error rate (BER) than AWGN channel, but lower than that of Rayleigh fading channel. For an example, while using the QAM modulation scheme, for SNR value 13, BER for Rician fading channel remains 7.5758e-07 while AWGN channel has zero BER. Again, for SNR value 14, both Rician fading channel and AWGN channel has zero BER while Rayleigh fading channel has BER value 2.1212e-05. After that, for SNR value 15 to 17, BER for Raleigh fading channel remains non-zero while BER for AWGN & Rician fading channels remain zero.

IV. CONCLUTION

In this research work, it has been studied the performance of an OFDM based WIMAX Communication system adopting different coding schemes and digital modulation scheme; M-ary QAM. A range of system performance results highlights the impact of AWGN and fading (Rayleigh & Rician) channels under QAM, 16-QAM & 64-QAM modulation techniques. From this research work, conclusions can be drawn regarding the BER performance evaluation of WIMAX Communication system over AWGN channel and fading (Rayleigh & Rician) channels like as below:

1. The performance of AWGN channel is the best of all channels as it has the lowest bit error rate (BER) under QAM, 16-QAM & 64-QAM modulation schemes. The amount of noise occurs in the BER of this channel is quite slighter than fading channels.

2. The performance of Rayleigh fading channel is the worst of all channels as BER of this channel has been much affected by noise under QAM, 16-QAM & 64-QAM modulation schemes.

3. The performance of Rician fading channel is worse than that of AWGN channel and better than that of Rayleigh fading channel. Because Rician fading channel has higher BER than AWGN channel and lower than Rayleigh fading channel. BER of this channel has not been much affected by noise under QAM, 16-QAM & 64-QAM modulation schemes.

REFERENCES

[1] "Nextel Flash-OFDM: The Best Network You May Never Use". PC Magazine. March 2, 2005. Retrieved July 23, 2011

[2] Raza Akbar, Syed Aqeel Raza, Usman Shafique, "PERFORMANCE EVALUATION OF WIMAX", Blekinge Institute of Technology, March 2009.

[3] Dennis Roddy, "Satellite Communcations," Third edition,McGraw-Hill Telecom Engineering.

[4] Theodore S. Rapaport, "Wireless Communications Principles and Practice," Prentice-Hall of India Private Limited,2004

[5] J. G. Proakis, Digital Communications, McGraw-Hill Inc., New York, NY, 1995 (Third Edition).

[6] Jingxin Chen, "CARRIER RECOVERY IN BURST-MODE 16-QAM", June 2004

http://sites.google.com/site/ijcsis/
ISSN 1947-5500

[7] WAN FARIZA BINTI PAIZI @ FAUZI , "BER PERFORMANCE STUDY OF PSK-BASED DIGITAL MODULATION SCHEMES IN MULTIPATH FADING ENVIRONMENT", JUNE 2006.

[8] Kaveh Pahlavan and Prashant Krishnamurthy, "Principles Of Wireless Networks",Prentice-Hall of India Private Limited, 2002.

http://sites.google.com/site/ijcsis/
ISSN 1947-5500

Rule Based Hindi to English Transliteration System for Proper Names

Monika Bhargava [#1], M.Kumar [*2], Sujoy Das [#3]

[# 1] *M.Tech Scholar CSE Department, SIRT, Bhopal, India*

[*2] *Professor CSE Department, SIRT, Bhopal, India*

[#3] *Associate Professor, Department of Computer Application, MANIT, Bhopal, India*

[1] monika_bhrgv@yahoo.co.in

[2] prof.mkumar@gmail.com

[3] sujdas@gmail.com

***Abstract*— There are Cross Language Information Retrieval systems that uses bilingual dictionary for translating user query from one language to another. The problem arises when a query term is not available in the bilingual dictionary. Such words are called Out of Vocabulary (OOV) words, and should be transliterated during translation process.OOV words are mainly proper nouns, named entity, and technical terms. We have developed a rule based transliteration system from Hindi to English script. We have also created a database of specialized spelling, e.g. some city names, person names, etc. which has considerably improved performance of our system.**

Keywords- CLIR, OOV, Transliteration

I. Introduction

In past 20 years the area of Information Retrieval (IR) has grown well beyond its primary goals of indexing text and searching useful documents in a collection. Nowadays, research in IR includes modelling, document classification and categorization, data visualization, filtering, etc. The Web is becoming a universal repository of human knowledge and culture which has allowed unprecedented sharing of ideas and information in a scale never seen before. Now the Web is seen as a publishing medium with accessibility to everybody. The Web contents are growing very rapidly and contain information written in many languages. Often a user of Web needs information written in a language not familiar to the user but he /she wishes to get it in the native language. This is possible through a process called Cross Language Information Retrieval. Several methods are used to convert the text of one language to another language. Machine translation systems in many language pairs are available and widely used. The bilingual dictionaries are also frequently used to convert text from one language to another language. A major problem arises when a word of the text is not available in the bilingual dictionary. Such words are called Out of Vocabulary (OOV) words and should be transliterated.

Transliteration is the task of transcribing a word or text from one writing system into another writing system such that pronunciation of the word remains same and a person reading the transcribed words can read it in original language. In others words, transliteration is the task of converting a text in its customary orthography. It is different from translation, which generates the meaning of the input text e.g. book is translated as किताब but transliterated as बुक. OOV words are problematic in Cross Lingual Information Retrieval. A common source of error in CLIR is out of vocabulary words, named entity and technical terms.

Among OOV, the proper nouns pose a major problem in the transliteration .This is due to the fact that a proper noun (name of person) is written by different persons with different spelling. This research has developed a rule based Hindi-English transliteration system especially for proper nouns with a fair degree of accuracy.

II. Literature Survey

The problem of transliteration has been studied by a number of researchers during the last decade. Knight and Graehl [1] use five probability distributions at various phases of transliteration for the language pair English to Katakana (a form of Japanese Language) writing system. Al-Onaizan and Knight [2] have studied transliteration system from Arabic to English writing, which uses existing named entity recognition system. Asif et.al [3] have considered Bengali to English transliteration scheme and used supervised training set to obtain a direct orthographic mapping. Lehal and Saini [4] have developed a Hindi to Urdu transliteration system by improving on the work of Bushra and Tafseer [5]. Lehal and Saini have claimed an accuracy of 99.46% when Hindi Unicode text is transliterated to Urdu.

Haung et.al [6] have developed a system which extracts Hindi – English named entity pairs through alignment of parallel corpus. Here, Chinese-English pairs are first extracted using a dynamic programming string matching. This model is then adapted for Hindi-English named entity pairs. Sinha et.al [7] have developed a simple yet powerful method for mining of Hindi – English names from parallel text corpus. The Hindi text written in Devanagari is first converted to IITK-Roman form which is direct representation of UTF-8 or ISCII -8 coding scheme and claimed an accuracy of nearly 93%.

In this paper, an effort is made to develop a rule based transliteration scheme for proper names of Hindi-English language pair. The system is under extensive experimentation and test.

http://sites.google.com/site/ijcsis/
ISSN 1947-5500

III. Overview of Hindi and English Scripts

Hindi, official language of India, is an Indo-Aryan language with about 487 million speakers. Hindi is written in Devanagri script which uses 52 symbols for representing 10 vowels, 40 consonants and 2 modifiers. The vowels are transcribed in two forms i.e. independent and dependent form. Dependent form is also known as matraa. Former is used when vowel letter appears alone at the beginning of word or is immediately followed by another vowel. Latter is used when vowel followed consonant [8].

English is the most widely used language in the world. Approximately 375 million people speak English. It has been referred to as 'world Language. English speaker have many different accents which often signal the speaker's native dialect or language. English is derived from West Germanic branch of Indo –European family. English has 21 consonants and 5 vowels.

The Indian languages TRANSliteration (ITRANS) is an ASCII transliteration scheme for Indic scripts, particularly for Devanagari script. [9]. It is a pre-processor that converts English-encoded text into various Indian languages script and has 7-bit ASCII encoding schemes (see [10]).

Mapping from Hindi to English

There is no one to one correspondence from Hindi to English script. Tables I and II show the ITRANS mapping between source language [Hindi] to target language [English].We have used these mappings to transliterate proper names in Hindi to English language.

TABLE I
MAPPING OF VOWELS FROM HINDI TO ENGLISH

S.No	Hindi Vowel		English Vowel
	Dependent Form	**Independent Form**	
1		अ	A
2	ा	आ	AA
3	ि	इ	I
4	ी	ई	II
5	ु	उ	U
6	ू	ऊ	UU
7	ृ	ऋ	RRI
8	े	ए	E
9	ै	ऐ	AI
10	ो	ओ	O
11	ौ	औ	AU
12	ं	अं	AM
13	ः	अः	AH

TABLE II
MAPPING OF MOST CONSONANTS FROM HINDI TO ENGLISH

S.No	Hindi	English
1	क	KA
2	ख	KHA
3	ग	GA
4	घ	GHA
5	ङ	NA
6	च	CHA
7	छ	CHHA
8	ज	JA
9	झ	JHA
10	ञ	NA
11	ट	TA
12	ठ	THA
13	ड	DA
14	ढ	DHA
15	ण	NA
16	त	TA
17	थ	THA
18	द	DA
19	ध	DHA
20	न	NA
21	प	PA
22	फ	PHA
23	ब	BA
24	भ	BHA
25	म	MA
26	य	YA
27	र	RA
28	ल	LA
29	व	VA OR WA
30	श	SHA
31	ष	SHA OR SHHA
32	स	SA
33	ह	HA
34	क्ष	KSHA
35	ज्ञ	GYA

http://sites.google.com/site/ijcsis/
ISSN 1947-5500

IV. Experimental Setup

We have developed a rule based transliteration system using JAVA SE DEVELOPMENT KIT (JDK), VERSION 6, MYSQL Server 5.5 for database. The program flow and the system architecture of Hindi to English rule based transliterator is shown in Fig1. The string of Hindi language is first searched in the database created for specialized spelling used in proper names, if the string match is found in the database then its transliteration equivalent is produced. If the string is not found in the database, it is then transliterated using nine rules given below:

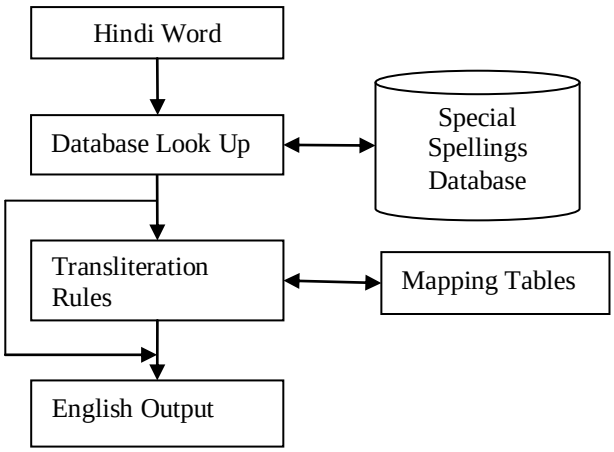


Fig. 1. Program Flow of the Transliteration system

Observations and Rules Creation

Sample results of our experimentation are shown in Tables III-XI where sample Hindi strings, expected English transliteration (commonly used English spellings), transliteration through ITRANS mappings and results produced by our transliteration system are shown.

TABLE III
SAMPLE 1 (TOTAL 45 STRINGS TESTED)

Hindi String	Expected Transliteration	Through Mapping	Rule Based Transliteration
पलक	PALAK	PALAKA	PALAK
अभय	ABHAY	ABHAYA	ABHAY
भरत	BHARAT	BHARATA	BHARAT
करण	KARAN	KARANA	KARAN

1) Rule1: The observation of Table III gives the rule that if length of proper noun is of 3 characters containing no vowel (matraa) then 'A' is removed from the last position.

TABLE IV
SAMPLE 2 (TOTAL 45 STRINGS TESTED)

Hindi String	Expected Transliteration	Through Mapping	Rule Based Transliteration
हरशद	HARSHAD	HARASHADA	HARSHAD
तनमय	TANMAY	TANAMAYA	TANMAY
अरनव	ARNAV	ARANAVA	ARNAV
अकबर	AKBAR	AKABARA	AKBAR

2) Rule2: The observation of Table IV gives the rule that if length of proper noun is of 4 characters containing no vowel (matraa) then 'A' is removed from the second and last position.

TABLE V
SAMPLE 3 (TOTAL 45 STRINGS TESTED)

Hindi String	Expected Transliteration	Through Mapping	Rule Based Transliteration
कपिल	KAPIL	KAPILA	KAPIL
अभिनव	ABHINAV	ABHINAVA	ABHINAV
विपिन	VIPIN	VIPINA	VIPIN
मुकुल	MUKUL	MUKULA	MUKUL

3) Rule3: The observation of Table V gives the rule that if proper noun ends with a consonant then 'A' should be removed from last position in English spelling.

TABLE VI
SAMPLE 4 (TOTAL 45 STRINGS TESTED)

Hindi String	Expected Transliteration	Through Mapping	Rule Based Transliteration
विमला	VIMLA	VIMALAA	VIMLA
पदमा	PADMA	PADAMAA	PADMA
बुशरा	BUSHRA	BUSHARAA	BUSHRA
बलदेव	BALDEV	BALADEVA	BALDEV

4) Rule4: The observation of Table VI gives the rule that if two consonants occur in succession, latter consonant followed by a vowel in a proper noun and index of first consonant should be greater than 1, then 'A' is removed from first consonant during transliteration.

TABLE VII
SAMPLE 5 (TOTAL 45 STRINGS TESTED)

Hindi String	Expected Transliteration	Through Mapping	Rule Based Transliteration
आसिफ	AASIF	AASIPHA	AASIF
आमना	AAMNA	AAMANAA	AAMNA
हिमानी	HIMANI	HIMAANII	HIMANI
गरिमा	GARIMA	GARIMAA	GARIMA

5) Rule5: The observation of Table VII gives the rule that if proper noun begins with 'आ' it is replaced with 'AA', and if

http://sites.google.com/site/ijcsis/
ISSN 1947-5500

diacritical form 'ा' occurs in between or end then it is replaced with 'A' during transliteration.

TABLE VIII
SAMPLE 6(TOTAL 45 STRINGS TESTED)

Hindi String	Expected Transliteration	Through Mapping	Rule Based Transliteration
ईश्वर	ISHWAR	IISHWARA	ISHWAR
करीना	KAREENA	KARIINAA	KAREENA
दीपक	DEEPAK	DIIPAKA	DEEPAK
मानसी	MANSI	MAANASII	MANSI

6) Rule6: The observation of Table VIII gives the rule that if proper noun begins with 'ई' or ends with its diacritical form 'ी' then it is replaced with 'I' and if its diacritical form occurs in middle then it is replaced with 'EE' instead of 'II' during transliteration.

TABLE IX
SAMPLE 7 (TOTAL 45 STRINGS TESTED)

Hindi String	Expected Transliteration	Through Mapping	Rule Based Transliteration
पूनम	POONAM	PUUNAMA	POONAM
अपूर्वा	APOORVA	APUURVAA	APOORVA
चारू	CHARU	CHAARUU	CHARU
कुहू	KUHU	KUHUU	KUHU

7) Rule7: The observation of Table IX gives the rule that if diacritical form of ऊ (ू) occurs in middle of proper noun then it is replaced with 'OO' and if its occurs in end of name it is replaced with 'U' instead of 'UU' during transliteration.

TABLE X
SAMPLE 8 (TOTAL 45 STRINGS TESTED)

Hindi String	Expected Transliteration	Through Mapping	Rule Based Transliteration
आफताब	AAFTAB	AAPHATAAB	AAFTAB
फाल्गुनी	FALGUNI PHALGUNI	PHAALGUNII	FALGUNI
आफरीन	AAFREEN	AAPHARIINA	AAFREEN
सैफ	SAIF	SAIPHA	SAIF

8) Rule8: The observation of Table X gives the rule that whenever a name contains 'फ' then mapping of 'फ' becomes 'F' instead of 'PH'.

TABLE XI
SAMPLE 9 (TOTAL 45 STRINGS TESTED)

Hindi String	Expected Transliteration	Through Mapping	Rule Based Transliteration
ओजस्वी	OJASWI	OJASVII	OJASWI
स्वपनिल	SWAPNIL	SVAPANILA	SWAPNIL
श्वेता	SHWETA	SHVETAA	SHWETA
विश्वा	VISHWA	VISHVAA	VISHWA

9) Rule9: The observation of Table XI gives the rule that whenever a name contains 'र' or 'श' followed by 'व' then the mapping of 'व' becomes 'W' instead of 'V'.

V. CONCLUSION

Many authors have remarked that the rule based transliteration system is very complex to develop (see [4, 5, and 7]).However, in this paper we have presented a Rule Based Transliteration system from Hindi to English for proper names. We have used standard ITRANs mapping (shown in Table 1 and 2) for our transliteration system. We have performed experimentation with 45 strings each of the similar types and produced nine rules for correcting the output, which matches with the expected (commonly used) spellings. We observed that there were several names, for example, city names, train names, etc, have specialized spellings. We have created a database containing specialized spellings. This has increased the performance of the system considerably. The transliteration system is under extensive test and some additional rules will be reported soon.

REFERENCES

[1] Knight K. and J. Graehl, "Machine Transliteration", *Computational Linguistics*, 24(4): pp 599-612, 1998.
[2] Al-Onaizan Y. and Knight K., "Translating Named Entities Using Monolingual and Bilingual Resources", *Proceedings of ACL 2002*, pp 400-408, July2002.
[3] Ekbal Asif, Sudip Kumar Naskar and Sivaji Bandyopadhyay, "A Modified Joint Source-Channel Model for Transliteration", *Proceedings of ACL 2006*, pp 191-198, 2006.
[4] Lehal G.S and Saini T.S., "A Hindi to Urdu Transliteration System", *Proceedings of ICON*, pp 235-240,2010
[5] Bushra and Tafseer, "Hindi to Urdu Conversion: Beyond Simple Transliteration", *Proceedings of the Conference on Language and Technology*, pp 24-31, 2009.
[6] Huang Fei, Stephan Vogel, and Alex Waibel, "Extracting Named Entity Translingual Equivalence with Limited Resources", *ACM Transactions on Asian Language Information Processing (TALIP)*,2(2):pp 124–129,2003
[7] R. Mahesh, K. Sinha, "Automated Mining Of Names Using Parallel Hindi-English Corpus", *7th Workshop on Asian Language Resources, ACLIJCNLP 2009*, pp 48–54, 2009.
[8] http://hindilanguage.info/devanagari/
[9] http://en.wikipedia.org/wiki/ITRANS
[10] http://www.aczoom.com/itrans

Fingerprint Hiding in True Color Image

Shahd Abdul-Rhman Hasso[1] Maha Abdul-Rhman Hasso[2] Omar Saad[3]

[1],[2],[3] *Department of Computer Science, College of Computer Sciences and Math., University of Mosul*
Mosul, Iraq

***Abstract*— With the progress of the times and the use of technology the importance of the use of biometric is appear to get to the regulations that require identification of persons before using to reduce the risk of security attack. One of the oldest, most prominent and commonly used of these biometrics is the fingerprint, the finger is containing features that have an impact on any surface touched, especially smooth ones, from this the fingerprint become an evidence that inferred by the people.**

In this work, an algorithm of hiding fingerprint in an image is proposed. the fingerprint is an image of (8 bit) hidden in an image of true color (24bit) form as 1 bit and 2 bits and 4 bits. This is done using the algorithm well known algorithm Least Significant bit (LSB), which helps to hide the images and messages in the less important bits in the original image, also the hiding is applied in one of the three true colors. There is no visually sense on the image before and after hiding, no changes occur on the image. The histogram also used to test the changes on an image.

Keywords-component; LSB; Biometric; Fingerprint; True color images.

I. INTRODUCTION

As a result of the development in all fields, many manipulation systems, which form a major threat to the stored information, more important than this information is the security of the information itself. Thus, we find scientists and specialists are working hard to protect them from unauthorized, and which reached to the use of biometrics of the same person to get to the systems that require identification before they are used to reduce the risk of security attacks [1].

the most prominent, oldest commonly used of these biometrics is the fingerprint, in 1885, William Herschel proved to the world that the fingerprint is quite different from person to person, other Germany researchers proved that the fingerprint remains constant and does not change from birth to death when he took the fingerprint of his right finger and returned after forty years and took second fingerprint to the same finger and compare it with the past and found it did not change [2].

The finger contains features have an impact on any surface touched, especially smooth ones, from here it became one of the evidence and unequivocal evidence that inferred by the people [2].

It is also used an alternative to passwords that are highly vulnerable because of the suffering that many of us use simple words it is easier to remember which lead to many risks to be easily guessed and knowledge and also used for issuing passports in addition to other uses [3].

II. BIOMETRIC FEATURES

The biometric features are a mechanism to identify where the person to be verified based on his physiological or behavioral characteristics. On this basis the biometric features will be classified into three sections [4]:

Section I: physiological features such as:
- Fingers
- hand geometry
- DNA
- Retina and iris of the eye
- outline the blood vessels

Section II: behavioral traits such as:
- Signature
- Handwriting
- The rhythm of movement of the hand in the use of the keyboard.

Section III combines physiological and behavioral attributes together, such as:
- Voice recognition
- Identify the impact of the human foot

III. PREPARE YOUR PAPER BEFORE STYLING

The fingerprint is one verses of God in His creation, where each person imprints of its own – identification card in his fingers -. The fingerprint is a prominent protrusion in the skin neighboring depressions, everyone has a distinctive form. It has proven that two people in the world cannot match their fingerprint even identical twins. These features have its impact on every object you touch; therefore, the fingerprint recognition refers to the verification mechanism of the matching two fingerprints of a human. It is considered as types of biometrics that is used in person identification and verification [5].

A. The fingerprint advantages:

- Difficult forgery and theft.
- The verification process does not cost time and do not require an employees to do it.
- It cannot be guessed or forgotten as passwords.
- Fixed and does not change with age and even if they removed the skin layer is composed belong to the same characteristics in the new skin.
- Easy integration into portable devices, doors and cars. It can use instead of keys.
- used in the system work instead of the identification cards to prove attendance forced the

http://sites.google.com/site/ijcsis/
ISSN 1947-5500

staff to attend on their own to have their fingerprints and prove their identities [3] [4].

B. The fingerprint disadvantages:

- Need a special device to capture the image.
- Depends on the direct touch of the device thus may lead to transmission of infectious diseases.
- Suffer from the problem of error if found dust grains on the surface of the reader may not allow an authorized person to enter.
- Suffer like any other biometric of penetrate the database and change the existing samples thus allow unauthorized to access the system [3] [4].

IV. THE STEGANOGRAPHY

The Steganography is the hiding of a message - (data) within another message (other data) in order to hide the existence of the first message for a specific purpose, the covered data that is used as a circumstance or a bowl to hide can be a multimedia file such as images, text, audio or video files, ... etc. Also may be an executable files for different programs(.exe), thus, the process to hide is needing two files, one called the cover, and the other is the substance to be hidden [6].

V. THE STEGANALYSIS

The process of trying to detect the presence of hidden information, read, or to change or delete is called Steganalysis. To success the Steganalysis it must be: [7]

First, discovering that there is hidden information.
Second, change, delete or just read it.

All of our technologies is to hide data in a way not raise suspicions, do not leave any marks or trace evidence of a change. For example, in the case of hiding in images, it must consider several factors, including: Do not use well known images, or images that anyone can get copies of them (such as Google images) that facilitate comparison of two the images. As well as taking into account does not change image or blur it as change the colors clearly. Therefore, it is advised not to hide a lot of data at the same image fear of change in a way of destroying the primary purpose of the use of technology, raising suspicion that mean failure of the process.

It is difficult to identify the hidden data if the program is unknown to the intruder, but unfortunately some programs hide information in a way have an impact works like a radio broadcast news password! Therefore, attention should be paid when choosing a program to be used in the process of this hiding.

The steganography idea is to put a message inside the cover to form the hidden target. This can be represented by the following equation [6] [7]:

Target = hidden message want to hide + cover + stego key

In general, the steganography methods can be divided into four basic methods:

A. text hiding

The text can extracted the hidden message in it, either in a way that the text be the first letter of each word represents the letter of the message or hidden in a grammatical or verbal.

B. voice hiding

The message could be hidden in an audio signal in time or frequency domain by one of the following methods:

- Low-bit coding: This method has a high input capacity, but vulnerable to discovery.
- Extending the spectrum: the message is inserted into the higher frequencies than necessary.
- Coverage of perception: this way is more of input capacity, but the most vulnerable to discovery and is inserted into the text of the message inside the regions of the human person cannot be aware of the traffic light.

C. video hiding

Hiding is derived from the images where the videos are only a set of successive images.

D. Image hiding

The most method that has been studied by researchers. [6] [7].

VI. STEGANOGRAPHY IN LEAST SIGNIFICANT BIT(LSB)

LSB is the most commonly used method in the applications, it required to enter one or more bits of the message to be hidden and replaced with a less important bits from the image [8]. The least significant bit is deciding who has the lowest value calculations (20 = 1) while the most important decision is deciding which has the largest value calculations (27 = 128). For example: If you have an image made up of pixels each of which has 24-bit (24-bit image), it can be extracted 3 bits of each byte of the image by 3-bit of the message to hide. Therefore, the image of size (1024 × 768) can be hidden a text of its size (294 × 912) bytes without notice with the eye.

Another example: Imagine that the letter "G" is to hide inside the image carrier of eight bytes binary representation is:

10010101 00001101 11001001 10010110
00001111 11001011 10011111 00010000

We know that the representation of the "G" is 01000111 binary, then we can get the character in the image replacement becomes the least significant bit representation is:

10010100 00001101 11001000 10010110
00001110 11001011 10011111 00010001

VII. THE COLOR CONCEPT IN THE DIGITAL IMAGES

The color form red, green, blue (RGB) is a color model which combines the lights red, green and blue with each other in different ways to generate a wide range of colors. The main objective of the RGB color model is a sense,

http://sites.google.com/site/ijcsis/
ISSN 1947-5500

generate and display the images in electronic devices, such as computer screens [8].

VIII. THE PROPOSED METHOD

The purpose of the proposed algorithm is to hide the fingerprint of an authorized person then add it to the database and sent through the true image colors (24 bits). The fingerprint image was converted to a gray image and resized to 64×64 pixels then hide it in the less important binary bits of a cover image.

The work is applied to hide the fingerprint in 1 LSBit, 2 LSBits and 4 LSBits. The proposed algorithm is applied on the Red, green and blue form of image.

The proposed algorithm could be summarized as:

1- The fingerprint image file was read from database [9].
2- The fingerprint image file was converted to gray form and resized either to 32 × 32 Pixel or 64 × 64 Pixel.
3- Convert the fingerprint file to binary form.
4- The cover image file is read, it is a true color (24 bit) image file.
5- Select one of the three colors of an image (red, green or blue) to hide data in it by change the LSB of it by the fingerprint data in it.
6- Analyze result visually and by histogram.

IX. THE RESULTS

The analysis of results of the work is based on the visual sense of image before and after hiding a fingerprint inside and analysis of the histogram of the image with true color (24 bit) before and after the masking image fingerprint.

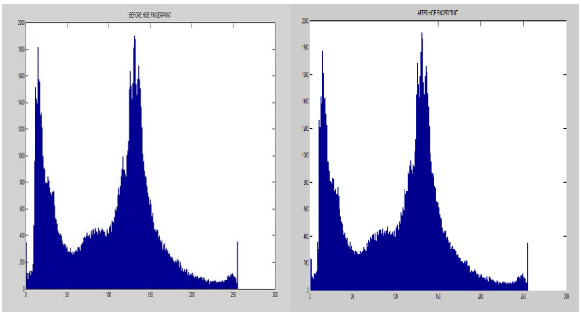
Figure (1) shows an image used in this work, the fingerprint image and the image after data hiding, where the figures (2) to (4) describes the histogram of some models of the cover before and after hiding so you will have three forms of histogram applied to image with the true colors before and after the hiding of one, two and 4-bit, respectively. The system has been applied on a set of models of color images and fingerprints we include a small part of them in figures (2) to (4).

Fingerprint Image

Image before hiding

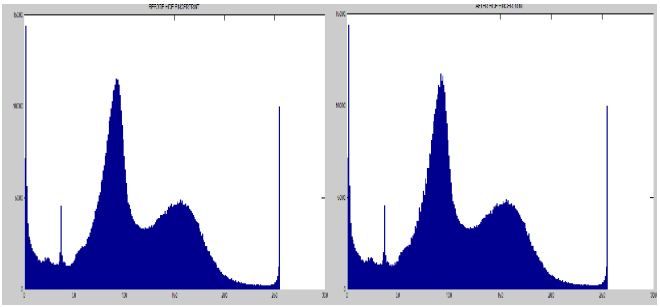
Image After hiding

Figure (1): A sample of an image for hiding fingerprint on it.



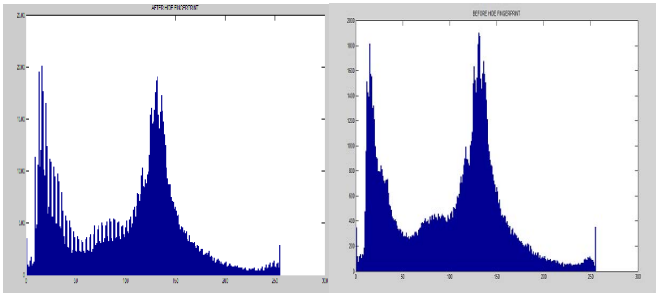
Histogram before hiding histogram after hiding

Figure (2): the histogram of an image before and after hiding data using one LSBit.



Histogram before hiding histogram after hiding

Figure (3): the histogram of an image before and after hiding data using two LSBits.



Histogram before hiding histogram after hiding

Figure (4): the histogram of an image before and after hiding data using four LSBits.

http://sites.google.com/site/ijcsis/
ISSN 1947-5500

X. CONCLUSION

The work explains that one could hide message (image) in a true color (24 bit) and hide the image regardless of the size of the image is done using known hiding algorithm to hide any biometric.

The applied algorithm in this work is LSB. It helps to hide images and messages in the form of true color image in one bit, two, or four bits of the message of the original image is no concealment in the last bit which is less important than the rest of the other bits.

This work has been hidden image of a fingerprint (8 bit) in the image of a true color (24 bit) in one bit, two and four bits. It is carefully that the hiding of one bit better than two and the two is better than four, i.e., in increasing of number of bits the hiding sense is increased and the security is decrease although in the proposed work the hiding is successful and no sense in the images after hiding.

RECOMMENDATIONS

- It can be encrypted fingerprint image by one of the encryption algorithms then hide in the image.
- It can be compress fingerprint image prior to hiding.
- It can be add another level of protection, a fingerprint examination of the person in the database before the addendum.

XI. REFERENCES

[1] Agrawal Neha, Savvides Marios, (2009), "Biometric Data Hiding: A 3 Factor Authentication Approach to Verify Identity with a Single Image Using Steganography, Encryption and Matching", IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops (CVPR Workshops).

[2] Potdar Vidyasagar M., Han Song, Chang Elizabeth, (2005), "Fingerprinted Secret Sharing Steganography for Robustness against Image Cropping Attacks", 3rd IEEE International Conference on Industrial Informatics (INDIN).

[3] Al-Khateeb Zina Nabeel, (2011) " Biometric identification based on Hand geometry", M Sc. Thesis, Department of Computer Science,Colloege of of Computer Sciences and Math., College, University of Mosul, Iraq

[4] Deswqi, Faiza Ahmed (2010)," Biometrics and information Security", Sixth conferense for library comettii anf information, Saudia Arbia.

[5] Steef Welan (2004), with the support of the employeers of the encyclopedia for helping poors CGAP and Echange LLC, " biometric Technologies" , series of improvement in information technology , by encyclopedia for helping poors.

[6] Rodriguez Benjamin M., Peterson Gilbert L., Agaian Sos S., (2007), "Multi-Class Classification Averaging Fusion for Detecting Steganography", IEEE International Conference on System of Systems Engineering, 2007, pp 309-314

[7] Mahmoud Hanan , Al-Hulaibah Hanan Saad , Al-Naeem Sarah Ahmad , Al-Qhatani Suha Ali , Al-Dawood Aljoharah , Al-Nassar Buthaina Saleh , AL-Salman Dhay Yousef , (2010), "Novel Technique for Steganography in Fingerprints Images: Design and Implementation", Sixth International Conference on Information Assurance and Security.

[8] Boukhonine Serguei, Krotov Vlad and Rupert Barry, (2005), "Future Security Approaches And Biometrics", Communications of the Association for Information Systems, Volume 16, pp., (937-966).

[9] Gonzales Rafael C. & Wood ,R.E ., (2008),"Digital image Processing" 3rd Edition , Publisher : Prentice.

[10] FVC2004 - Third International Fingerprint Verification Competition; FingerPrint Databases

AUTHOR PROFILE

Mrs. Shahd A. R. Hasso (M Sc.) is currently a lecturer at Mosul University/ College of Computer Science and Mathematics/ Computer Science Department. She received B.Sc. degree in Computer Science from University of Mosul in 1998 and M.Sc. degree from University of Mosul in 2003. Her research interests and activity are in data security, data structures, network security, information hiding. Now, she teaches data security undergraduate students.

Miss Maha A. R. Hasso (Ph. D.) is currently a lecturer at Mosul University/ College of Computer Science and Mathematics/ Computer Science Department. She received B.Sc. degree in Computer Science from University of Mosul in 1991, M.Sc. degree from University of Mosul in 1998 and Ph. D. degree from University of Mosul. Her research interests and activity are in image processing, computer vision, pattern recognition, remote sensing applications and biometrics. Now, she teaches digital image processing, pattern recognition and visual programming for postgraduate and undergraduate students.

Network Intrusion Detection Using Improved Decision Tree Algorithm

K.V.R. Swamy
Department Of Computer Science and Engineering
V.R. Siddhartha Engineering College
Vijayawada, Andhra Pradesh, India
swamy.vrsec@gmail.com

K.S. Vijaya Lakshmi
Department Of Computer Science and Engineering
V.R. Siddhartha Engineering College
Vijayawada, Andhra Pradesh, India
ksvl66@yahoo.co.in

***Abstract* – Intrusion detection involves a lot of tools that are used to identify different types of attacks against computer systems and networks. With the development of network technologies and applications network attacks are greatly increasing both in number and severe. Open source and commercial network intrusion detection tools are not able to predict new type of attacks based on the previous attacks. So, data mining is one of the methods used in IDS (Intrusion Detection System). In recent years data mining based network intrusion detection system has been giving high accuracy and good detection on different types of attacks. In this paper, the performance of the data mining algorithms like C4.5 and improved C4.5 are being used in order to detect the different types of attacks with high accuracy and less error prone.**

Keywords- C4.5 Decision Tree; Improved C4.5 Decision Tree; Intrusion detection system.

I. INTRODUCTION

Nowadays, many organizations and companies use Internet services as their communication and marketplace to do business such as at EBay and Amazon.com website. Together with the growth of computer network activities, the growing rate of network attacks has been advancing, impacting to the availability, confidentiality, and integrity of critical information data. Therefore a network system must use one or more security tools such as firewall, antivirus, IDS and Honey Pot to prevent important data from criminal enterprises.

A network system using a firewall only is not enough to prevent networks from all attack types. The firewall cannot defense the network against intrusion attempts during the opening port. Hence a Real-Time Intrusion Detection System (RT-IDS), shown in Fig 1, is a prevention tool that gives an alarm signal to the computer user or network administrator for antagonistic activity on the opening session, by inspecting hazardous network activities [1].

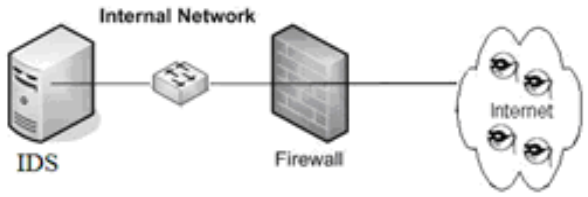


Fig 1: Intrusion detection system environment

IDSs have gained acceptance as a necessary addition to every organization's security infrastructure despite the documented contributions intrusion detection technologies make to system security, in many organizations one must still justify the acquisition of IDSs. We may use IDSs to prevent problem behaviors by increasing the perceived risk of discovery of those who would attack or abuse the system.

There are two general categories of attacks which intrusion detection technologies attempt to identify - anomaly detection and misuse detection. Anomaly detection identifies activities that vary from established patterns for users, or groups of users. Anomaly detection typically involves the creation of knowledge bases that contain the profiles of the monitored activities. The second general approach to intrusion detection is misuse detection. This technique involves the comparison of a user's activities with the known behaviors of attackers attempting to penetrate a system. While anomaly detection typically utilizes threshold monitoring to indicate when a certain established metric has been reached, misuse detection techniques frequently utilize a rule-based approach. When applied to misuse detection, the rules become scenarios for network attacks. The intrusion detection mechanism identifies a potential attack if a user's activities are found to be consistent with the established rules. The use of comprehensive rules is critical in the application of expert systems for intrusion detection [2].

http://sites.google.com/site/ijcsis/
ISSN 1947-5500

There are many methods applied into intrusion detection, such as methods based on statistics, methods based on data mining, methods based on machine learning and so on. In recent years, data mining technology is developing rapidly and increasingly mature. Now it is gradually applied to the intrusion detection field, and has made a number of important achievements at home and abroad. The basic principles of intrusion detection based on data mining are as follows: Firstly intelligently analyze and deal with security audit data from different data sources(such as host-based, network-based, alarm-based), this can help system generate intrusion rules and establish anomaly detection model by extracting regularity of data; Then use these knowledge to discriminate new network behaviors. The main methods are: classification analysis, clustering analysis, genetic algorithm, neural networks, association rule mining, sequential pattern mining, and outlier detection and so on. Decision tree technology is an intuitionist and straightforward classification method. It has great advantage in extracting features and rules. Therefore applying decision tree technology into intrusion detection is of great significance [3].

Locations of Intrusion Detection Systems in Networks:

Usually an intrusion detection system captures data from the network and applies its rules to that data or detects anomalies in it. Depending upon the network topology, the type of intrusion activity (i.e. internal, external or both), and our security policy (what we want to protect from hackers), IDSs can be positioned at one or more places in the network . For example, if we want to detect only external intrusion activities, and we have only one router connecting to the Internet, the best place for an intrusion detection system may be just inside the router or a firewall. On the other hand, if we have multiple paths to the internet, and we want to detect internal threats as well, we should place one IDS box in every network segment. Fig. shows typical locations where you can place an intrusion detection system.

II. RELATED WORK

In his paper, Except for the information gain measure and its improved versions, Lopez de Mantaras[4] presented a distance-based attribute selection measure. His experimental study proves that the distance based measure is not biased toward attributes with large numbers of values, and avoids the practical issues towards the gain ratio measure. Mingers[5] provides an experimental study of the relative accuracy of different attribute selection measures in the decision tree in order to overcome the bias in the tuples. Nageswara Rao, Dr. D. Rajya Lakshmi, Prof T. Venkateswara Rao et at[6] proposed robust statistical preprocessor in order to improve the accuracy. But the limitation in that paper is existing c45 does not handle when the dataset is large. An expert system consists of a set of rules that encode the knowledge of a human "expert". These rules are used by the system to make conclusions about the security-related data from the intrusion detection system. Expert systems permit the incorporation of an extensive amount of human experience into a computer application that then utilizes that knowledge to identify activities that match the defined characteristics of misuse and attack. Unfortunately, expert systems require frequent updates to remain current. While expert systems offer an enhanced ability to review audit data, the required updates may be ignored or performed infrequently by the administrator. At a minimum, this leads to an expert system with reduced capabilities. At worst, this lack of maintenance will degrade the security of the entire system by causing the system's users to be misled into believing that the system is secure, even as one of the key components becomes increasingly ineffective over time. Rule-based systems suffer from an inability to detect attacks scenarios that may occur over an extended period of time. Slight variations in an attack sequence can affect the activity-rule comparison to a degree that the intrusion is not detected by the intrusion detection mechanism. While increasing the level of abstraction of the rule-base does provide a partial solution to this weakness, it also reduces the granularity of the intrusion detection device[11].

III. PROPOSED ARCHITECTURE

Following framework gives the overall description about the proposed approach. In this framework,KDD dataset[7] is used as training data for classification purpose.

Proposed framework has following algorithms.

1) Min Max Normalization

2) Decision Tree Algorithms.

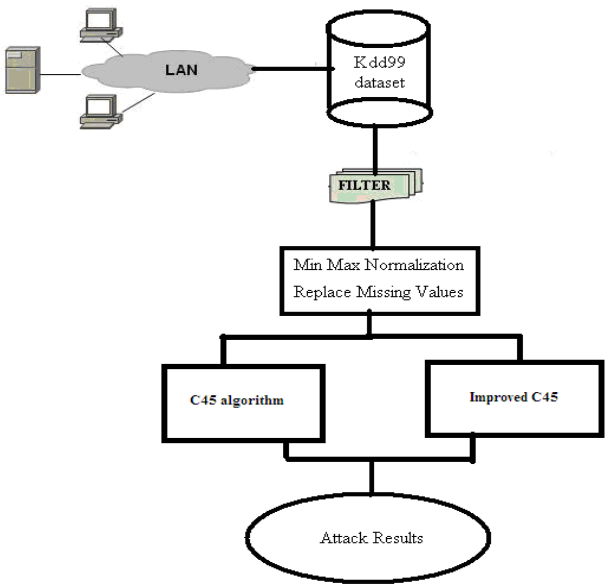


Fig 2: Proposed Framework

http://sites.google.com/site/ijcsis/
ISSN 1947-5500

A. KDD Dataset

The KDD Cup 1999 dataset was derived from the 1998 DARPA Intrusion detection evaluation program prepared and managed by MIT Lincoln Laboratory. The dataset was a collection of simulated raw TCP dump data over a period of nine weeks. There are 4,898,430 labeled and 311,029 unlabeled connection records in the dataset [8]. The labeled connection records consist of 41 attributes: 7 symbolic and 34 numeric. The complete listing of the set of features in the dataset is given in Table 1.

TABLE I: List of attributes in KDD dataset

No	Name of the attribute	No	Name of the attribute
1	duration	22	is_guest_login
2	protocol_type	23	count
3	service	24	srv_count
4	flag	25	serror_rate
5	src_bytes	26	srv_serror_rate
6	dst_bytes	27	rerror_rate
7	land	28	srv_serror_rate
8	wrong_fragment	29	same_srv_rate
9	urgent	30	diff_srv_rate
10	hot	31	srv_diff_host_rate
11	num_failed_logins	32	dst_host_count
12	logged_in	33	dst_host_srv_count
13	num_compromised	34	dst_host_same_srv_rate
14	root_shell	35	dst_host_diff_srv-rate
15	su_attempted	36	dst_host_same_srv_port_rate
16	num_root	37	dst_host_srv_diff_host_rate
17	num_file_creations	38	dst_host_serror_rate
18	num_shells	39	dst_host_srv_serror_rate
19	num_access_files	40	dst_host_rerror_rate
20	num_outbound_cmd	41	dst_host_srv_rerror_rate
21	is_host_login		

B. Data Transformation

In data transformation, the data are transformed or consolidated into forms appropriate for mining. Min-max normalization performs a linear transformation on the original data. Suppose that minA and maxA are the minimum and maximum values of an attribute A. Min-max normalization maps a value, v, of A to v0 in the range [new_minA, new_maxA] by computing

$$v' = \frac{v - min_A}{max_A - min_A}(new\_max_A - new\_min_A) + new\_min_A.$$

Min-max normalization preserves the relationships among the original data values. It will encounter an "out-of-bounds" error if a future input case for normalization falls outside of the original data range for A.

C. Data Preprocessing

Incomplete, noisy, and inconsistent data are commonplace properties of large real world databases and data warehouses. Incomplete data can occur for a number of reasons. Attributes of interest may not always be available. Other data may not be included simply because it was not considered important at the time of entry. Relevant data may not be recorded due to a misunderstanding, or because of equipment malfunctions. Data that were inconsistent with other recorded data may have been deleted. Furthermore, the recording of the history or modifications to the data may have been overlooked. Missing data, particularly for tuples with missing values for some attributes, may need to be inferred. There are many possible reasons for noisy data (having incorrect attribute values). Data cleaning (or data cleansing) routines attempt to fill in missing values, smooth out noise while identifying outliers, and correct inconsistencies in the data. Handling Missing Values: The attribute mean or stddev to fill in the missing value.

D. C45 ALGORITHM

Algorithm**:** Geneate_decision_tree

Input**:** Data partition, D, which is a set of training tuples and their associated class labels. Attribute_list, the set of candidate attributes. Attribute_selection_method, a procedure to determine the splitting criterion that "best" partitions the data tuples into individual classes. This criterion consists of a splitting_attribute and, possibly, either a split point or splitting subset.

Output**:** a decision tree

Method:

(1) create a node N;

(2) if tuples in D are all of the same class, C then

(3) return N as a leaf nod labeled with the class C;

(4) If attribute_list is empty then

http://sites.google.com/site/ijcsis/
ISSN 1947-5500

(5) Return N as a leaf node labeled with the majority class in D; //majority voting
(6) Apply attribute_seletion_method (D, arrtibute_list) to find the "best" splitting_criterion;
(7)Label node N with splitting_criterion;
(8)If splitting_attribute is discrete-valued and
Multiway splits allowed then // not restricted to binary trees
(9) attribute_list→attribute_list - splitting_attribute; //remove splitting_attribute
(10) for each outcome j of splitting_criterion // partition the tuples and grow sub-tees for each partition
(11) Let Dj be the set of a data tuples in D satisfying outcome j; // a partition
(12) If Dj is empty then
(13) Attach a leaf labeled with the majority class in D to node N;
(15) Else attach the node returned by Geneate_decision_tree (Dj, attribute list) to node N;
(16) Return N;

E. IMPROVED C45

(1) create a node N;
(2)if tuples in D are all of the same class, C then
(3) return N as a leaf node labeled with the class C;
(4) if attribute list is empty then
(5) return N as a leaf node labeled with the majority class in D; // majority voting
(6) apply Attribute selection to each attribute(L, attribute list) to find the "best" splitting criterion;
Gain measures how well a given attribute separates training examples into targeted classes. The one with the highest information is selected. Given a collection S of c outcomes The expected information needed to classify a tuple in D is given by

Modified Information or entropy is given as

ModInfo(D) $= -S_i \sum_{i=1}^{m} l \log \sqrt{S_i}$,m different classes

ModInfo(D) $= -S_i \sum_{i=1}^{2} l \log \sqrt{S_i}$

$= -S_1 \log \sqrt{S_1} - S_2 \log \sqrt{S_2}$

Where S_1 indicates set of samples which belongs to target class 'anamoly', S_2 indicates set of samples which belongs to target class 'normal'.

Information or Entropy to each attribute is calculated using

$$Info_A(D) = \sum_{i=1}^{v} |D_i| / |D| \times ModInfo(D_i)$$

The term Di /D acts as the weight of the jth partition. ModInfo(D) is the expected information required to classify a tuple from D based on the partitioning by A.

Information gain is defined as the difference between the original information requirement) and the new requirement .That is,

$$Gain(A) = Mod \inf o(D) - \inf o_A(D)$$

(7)Label node N with splitting_criterion;
(8)If splitting_attribute is discrete-valued and
Multiway splits allowed then // not restricted to binary trees
(9) attribute_list→attribute_list - splitting_attribute; //remove splitting_attribute
(10) for each outcome j of splitting_criterion // partition the tuples and grow sub-tees for each partition
(11) Let Dj be the set of a data tuples in D satisfying outcome j; // a partition
(12) If Dj is empty then
(13) Attach a leaf labeled with the majority class in D to node N;
(15) Else attach the node returned by Geneate_decision_tree (Dj, attribute list) to node N;
(16) Return N;

IV. EXPERIMENTAL RESULTS

All experiments were performed in a one-year-old computer with the configurations Intel(R) Core(TM)2 CPU 2.13GHz, 2 GB RAM, and the operation system platform is Microsoft Windows XP Professional (SP2). The dataset to be used in our experiments in KDD99 labeled dataset. The main reason we use this dataset is that we need relevant data that can easily be shared with other researchers, allowing all kinds of techniques to be easily compared in the same baseline. The common practice in intrusion detection to claim good performance with "live data" makes it difficult to verify and improve pervious research results, as the traffic is never quantified or released for privacy concerns. As our test dataset, the KDD99 dataset contains one type of normal data and 24 different types of attacks. For implementation Netbeans is used.

http://sites.google.com/site/ijcsis/
ISSN 1947-5500

The input is KDD data set. It is about 10% of KDD dataset.

	A	B	C	D	E	F	G	H	I	J	K	L
1	duration	protocol_t	service	flag	src_bytes	dst_bytes	land	wrong_fra	urgent	hot	num_faile	logged_in
2	0	tcp	ftp_data	SF	491	0	0	0	0	0	0	0
3	0	udp	other	SF	146	0	0	0	0	0	0	0
4	0	tcp	private	S0	0	0	0	0	0	0	0	0
5	0	tcp	http	SF	232	8153	0	0	0	0	0	1
6	0	tcp	http	SF	199	420	0	0	0	0	0	1
7	0	tcp	private	REJ	0	0	0	0	0	0	0	0
8	0	tcp	private	S0	0	0	0	0	0	0	0	0
9	0	tcp	private	S0	0	0	0	0	0	0	0	0
10	0	tcp	remote_jo	S0	0	0	0	0	0	0	0	0
11	0	tcp	private	S0	0	0	0	0	0	0	0	0
12	0	tcp	private	REJ	0	0	0	0	0	0	0	0
13	0	tcp	private	S0	0	0	0	0	0	0	0	0
14	0	tcp	http	SF	287	2251	0	0	0	0	0	1
15	0	tcp	ftp_data	SF	334	0	0	0	0	0	0	1
16	0	tcp	name	S0	0	0	0	0	0	0	0	0
17	0	tcp	netbios_n	S0	0	0	0	0	0	0	0	0
18	0	tcp	http	SF	300	13788	0	0	0	0	0	1
19	0	icmp	eco_i	SF	18	0	0	0	0	0	0	0
20	0	tcp	http	SF	233	616	0	0	0	0	0	1
21	0	tcp	http	SF	343	1178	0	0	0	0	0	1
22	0	tcp	mtp	S0	0	0	0	0	0	0	0	0
23	0	tcp	private	S0	0	0	0	0	0	0	0	0

Fig 3: KDD Dataset

The existing C4.5 decision tree gives the 95.7 percent of accuracy for detecting attacks.

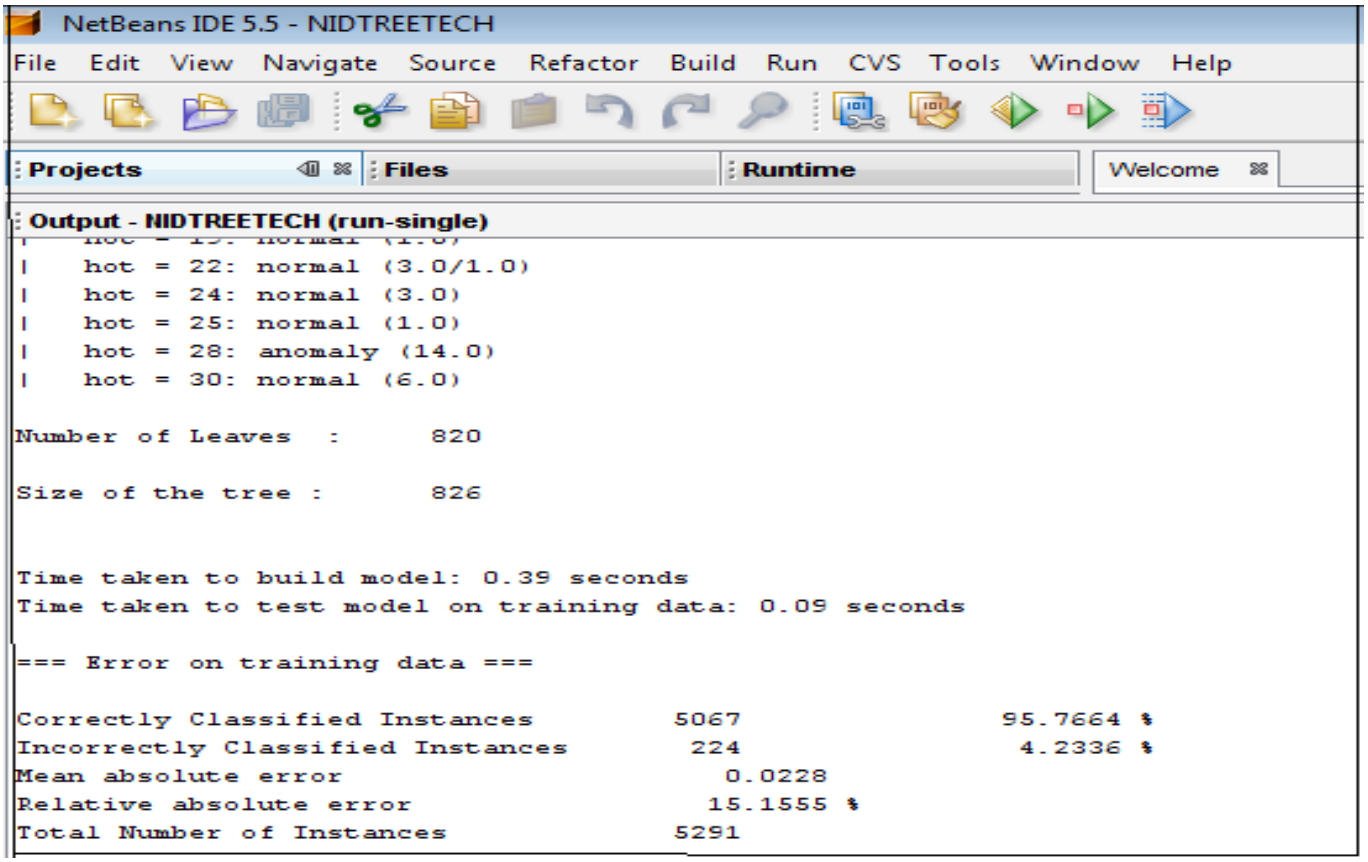


Fig 4: C45 decision tree result

http://sites.google.com/site/ijcsis/
ISSN 1947-5500

The proposed C4.5 decision tree gives the 96.7 percent of accuracy for detecting attacks with with less false positive and true negative rates.

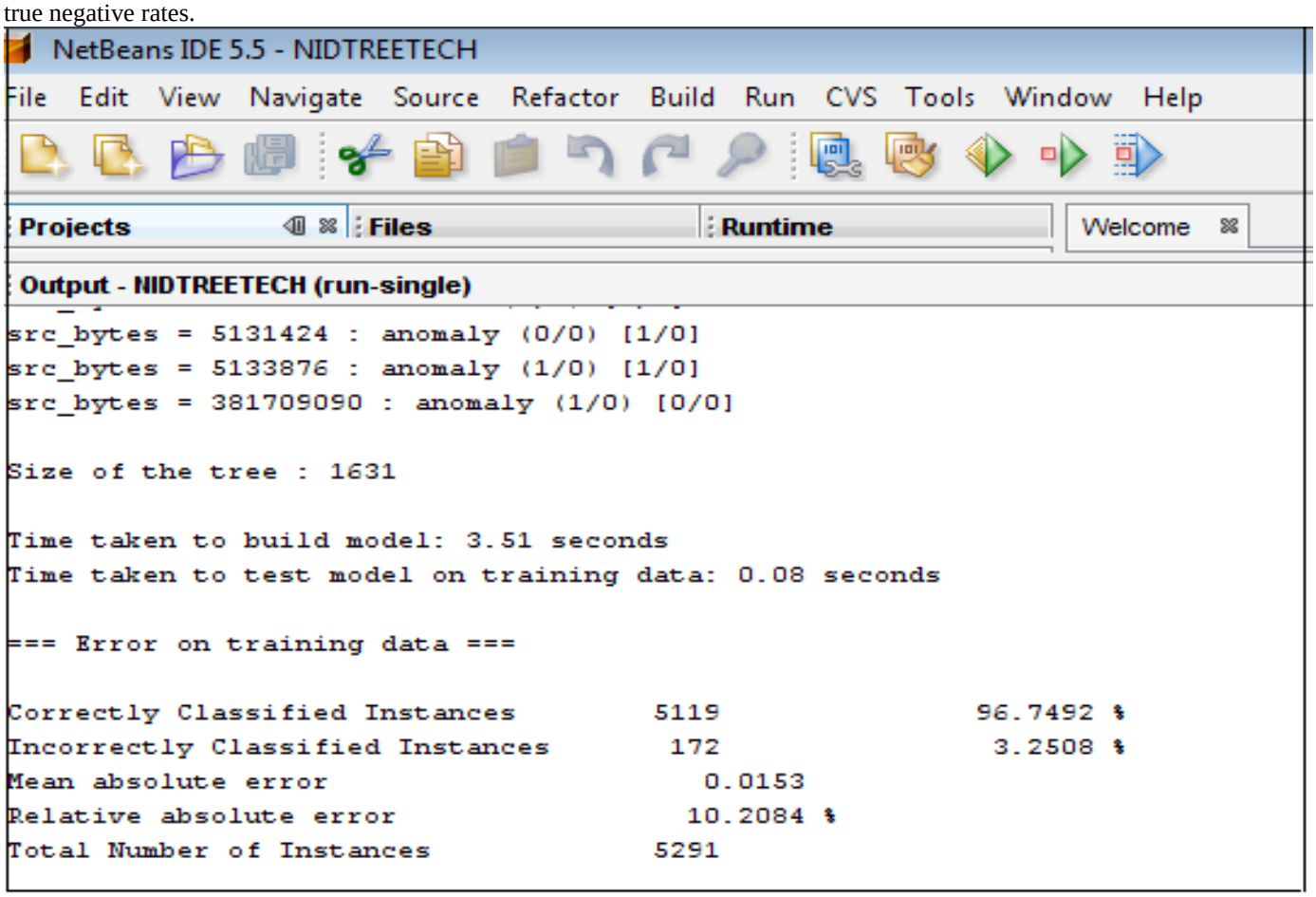


Fig 5: Improved C4.5 decision tree result

Following results gives the improved C45 performance on 10% KDD dataset with 5291 instances:

TABLE2: Improved C45 performance on 10% KDD dataset

PROPERTY	EXISTING C4.5	IMPROVED C4.5
Correctly Classified Instances	5067(95.76%)	5119(96.75%)
Incorrectly Classified Instances	224(4.23%)	172(3.25%)

http://sites.google.com/site/ijcsis/
ISSN 1947-5500

V. CONCLUSION

Experimental results show the existing C4.5 decision tree gives 95.7 percent of accuracy for detecting attacks. But the proposed decision tree gives better attack classified results compare to existing C4.5 technique. Proposed Algorithm gives 96.9 percent of accuracy for detecting attacks with less false positive and true negative rates. Data mining algorithms require an offline training phase, but the testing phase requires much less time and future work could investigate how well it can be adapted to performing online.

REFERENCES

[1] Real-time Intrusion Detection and Classification by Phurivit Sangkatsanee1, Naruemon Wattanapongsakorn and Chalermpol Charnsripinyo.

[2] Intelligent Adaptive Intrusion Detection Systems Using Neural Networks (Comparitive study) by Aida O. Ali, Ahmed I. saleh and Tamer R. Badawy.

[3] An intrusion detection algorithm based on decision tree technology by Juan Wang, Qiren Yang and Dasen Ren.

[4] R. L. de Mantaras "A distance-based attribute selection measure for decision tree induction. Machine Learning, 6:81–92, 1991
.

[5] J. Mingers "An empirical comparison of selection measures for decision-tree induction. Machine Learning, 3:319–342, 1989.

[6] Nageswararao,Dr.D.RajyaLakshmi,Prof T.Venkateswara Rao, " Robust Statistical Outlier based Feature Selection Technique for Network Intrusion Detection" ,(IJSCE 2012).

[7] Mahbod Tavallaee, Ebrahim Bagheri, Wei Lu, and Ali A. Ghorbani "A Detailed Analysis of the KDD CUP 99 Data Set", IEEE 2009.

[8] *www.cs.waikato.ac.nz/ml/**weka***

[9] J. R. Quinlan, "C4.5: programs for machine learning", Morgan Kaufmann Publishers, 1993.

[10] Hybrid Neural Network and C4.5 for Misuse Detection Zhi-Song Pan, Song-Can Chen, Gen-Bao Hu, Dao-Qiang Zhang, Proceedings of the Second International Conference on Machine Learning and Cybernetics, Xi"an, 2-5 November 2003.

[11] C. Kruegel, D. Mutz, W. Robertson, F. Valeur, "Bayesian event classification for intrusion detection," in Proc. of the 19th Annual Computer Security Applications Conference, Las Vegas, NV, 2003.

http://sites.google.com/site/ijcsis/
ISSN 1947-5500

Phases vs. Levels using Decision Trees for Intrusion Detection Systems

Heba Ezzat Ibrahim, Sherif M. Badr and Mohamed A. Shaheen
College of Computing and Information Technology
Arab Academy for Science, Technology and Maritime Transport
Cairo, Egypt
Heba_ezzat_86@yahoo.com

***Abstract*— Security of computers and the networks that connect them is increasingly becoming of great significance. Intrusion detection system is one of the security defense tools for computer networks. This paper compares two different model Approaches for representing intrusion detection system by using decision tree techniques. These approaches are Phase-model approach and Level-model approach. Each model is implemented by using two techniques, New Attacks and Data partitioning techniques. The experimental results showed that Phase approach has higher classification rate in both New Attacks and Data Partitioning techniques than Level approach.**

Keywords-component; network intrusion detection; Decision Tree; NSL-KDD dataset; network security

I. INTRODUCTION

The Internet and online procedures is an essential tool of our daily life. They have been used as a main component of business operation [1]. Therefore, network security needs to be carefully concerned to provide secure information channels [2].

It is difficult to prevent attacks only by passive security policies, firewall, or other mechanisms. Intrusion Detection Systems (IDS) have become a critical technology to help protect these systems as an active way. An IDS can collect system and network activity data, and analyze those gathered information to determine whether there is an attack [3].

Network Intrusion detection (NIDS) and prevention systems (NIPS) serve a critical role in detecting and dropping malicious or unwanted network traffic [5]. Intrusion detection and prevention systems (IDPS) are primarily focused on identifying possible incidents, logging information about them, attempting to stop them, and reporting them to security administrators. In addition, organizations use IDPSs for other purposes, such as identifying problems with security policies, documenting existing threats, and deterring individuals from violating security policies. IDPSs have become a necessary addition to the security infrastructure of nearly every organization [6].

Intrusion detection started in around 1980s after the influential paper from Anderson [4]. Intrusion detection systems are classified as network based, host based, or application based depending on their mode of deployment and data used for analysis [7]. Additionally, intrusion detection systems can also be classified as signature based or anomaly based depending upon the attack detection method. The signature-based systems are trained by extracting specific patterns (or signatures) from previously known attacks while the anomaly-based systems learn from the normal data collected when there is no anomalous activity [7].

Another approach for detecting intrusions is to consider both the normal and the known anomalous patterns for training a system and then performing classification on the test data. Such a system incorporates the advantages of both the signature-based and the anomaly-based systems and is known as the Hybrid System. Hybrid systems can be very efficient, subject to the classification method used, and can also be used to label unseen or new instances as they assign one of the known classes to every test instance. This is possible because during training the system learns features from all the classes. The only concern with the hybrid method is the availability of labeled data. However, data requirement is also a concern for the signature-based and the anomaly-based systems as they require completely anomalous and attack free data, respectively, which are not easy to ensure [8].

II. PREVIOUS WORK

The purpose of IDS is to help computer systems with how to discover attacks, and that IDS is collecting information from several different sources within the computer systems and networks and compares this information with preexisting patterns of discrimination as to whether there are attacks or weaknesses [10].

Decision Trees (DT) have also been used for intrusion detection [11]. Decision Tree is very powerful and popular machine learning algorithm for decision-making and classification problems. It has been used in many real life applications like medical diagnosis, radar signal classification, weather prediction, credit approval, and fraud detection etc

http://sites.google.com/site/ijcsis/
ISSN 1947-5500

[12]. The decision tree is a simple if then else rules but it is a very powerful classifier and proved to have a high detection rate. They are used to classify data with common attributes. Each decision tree represents a rule which categorizes data according to these attributes. A decision tree has three main components: nodes, leaves, and edges. Each decision tree represents a rule set, which categorizes data according to the attributes of dataset. The DT building algorithms may initially build the tree and then prune it for more effective classification. [13].

A. C5.0 Decision Trees

See5.0 (C5.0) is one of the most popular inductive learning tools originally proposed by J.R.Quinlan as C4.5 algorithm (Quinlan, 1993) [13].

C5.0 can deal with missing attributes by giving the missing attribute the value that is most common for other instances at the same node. Or, the algorithm could make probabilistic calculations based on other instances to assign the value [14].

B. Classification and Regression Trees (CRT or CART)

CART is a recursive partitioning method to be used both for regression and classification. The key elements of CART analysis are a set of rules for splitting each node in a tree; deciding when tree is complete and assigning a class outcome to each terminal node. CART is constructed by splitting subsets of the data set using all predictor variables to create two child nodes repeatedly, beginning with the entire data set [15].

C. Chi-squared Automatic Interaction Detector (CHAID)

CHAID (Chisquare-Automatic-Interaction-Detection) was originally designed to handle nominal attributes only.
CHAID method is based on the chi-square test of association. A CHAID tree is a decision tree that is constructed by repeatedly splitting subsets of the space into two or more child nodes, beginning with the entire data set [16].
CHAID handles missing values by treating them all as a single valid category. CHAD does not perform pruning.

D. Quick, Unbiased, Efficient Statistical Tree (QUEST)

QUEST is a binary-split decision tree algorithm for classification and machine learning. QUEST can be used with univariate or linear combination splits. A unique feature is that its attribute selection method has negligible bias. If all the attributes are uninformative with respect to the class attribute, then each has approximately the same change of being selected to split a node [17].

We compare between the phase model in [9], and the Level model in [6].The authors in [9] design a system which consists of three detection levels. The network data are introduced to the module of the first level which aims to differentiate between normal and attack. If the input record was identified as an attack then the administrator would be alarmed that the coming record is suspicious and then this suspicious record would be introduced to the second level which specifies the class of this attack (DOS, probe, R2L or U2R). The third detection level consists of four modules one module for each class type to identify attacks of this class. Finally the administrator would be alarmed of the expected attack type.
In [6], the authors classify network intruders into a set of different levels. The first level is called the Boolean detection level, where the system classifies the network users to either normal or intruder. The second level is called the coarse detection level, where it can identify four categories of intruders. The third level is called the fine detection level, where the intruder types can be fine tuned into 23 intruder types.

III. System Architecture

The system components :

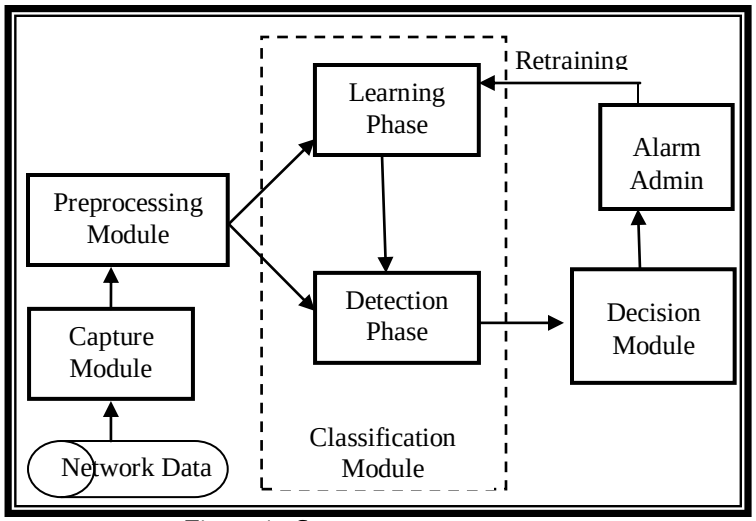


Figure 1. System components

Figure 1. shows the main modules of IDS as follows:

A. The Capture Module

Raw data of the network are captured and stored using the network adapter. It utilizes the capabilities of the TCP dump capture utility for Windows to gather historical network packets.

B. The Preprocessing Module

The data must be of uniform representation to be processed by the classification module. The preprocessing module is responsible for reading, processing, and filtering the audit data to be used by the classification module. The preprocessing module handles Numerical Representation, Normalization and Features selection of raw input data. The preprocessing module consists of three phases: [18]

1) Numerical Representation: Converts non-numeric features into a standardized numeric representation. This process involved the creation of relational tables for each of the data type and assigning a number to each unique type of

http://sites.google.com/site/ijcsis/
ISSN 1947-5500

element. (e.g. protocol_type feature is encoded according to IP protocol field: TCP=0, UDP=1, ICMP=2). This is achieved by creating a transformation table containing each text/string feature and its corresponding numeric value.

2) Normalization: The ranges of the features were different and this made them incomparable. Some of the features had binary values where some others had a continuous numerical range (such as duration of connection). As a result, inputs to the classification module should be scaled to fall between zero and one [0, 1] range for each feature.[9]

3) Dimension reduction: reduce the dimensionality of input features of the classification module. Reducing the input dimensionality will reduce the complexity of the classification module, and hence the training time.

C. The classification Module

The classification module has two phases of operation. The learning and the detection phase.

1) The Learning Phase

In the learning phase, the classifier uses the preprocessed captured network user profiles as input training patterns. This phase continues until a satisfactory correct classification rate is obtained.

2) The Detection Phase

Once the classifier is learned, its capability of generalization to correctly identify the different types of users should be utilized to detect intruder. This detection process can be viewed as a classification of input patterns to either normal or attack.

D. The Decision Module

The basic responsibility of the decision module is to transmit an alert to the system administrator informing him of coming attack. This gives the system administrator the ability to monitor the progress of the detection module.

To evaluate our system we used two major indices of performance. We calculate the detection rate and the false alarm rate according to the following assumptions [19]:

- False Positive (FP): the total number of normal records that are classified as anomalous
- False Negative (FN): the total number of anomalous records that are classified as normal
- Total Normal (TN): the total number of normal records
- Total Attack (TA): the total number of attack records
- Detection Rate = [(TA-FN) / TA]*100
- False Alarm Rate = [FP/TN]*100
- Correct Classification Rate = Number of Records Correctly Classified / Total Number of records in the used dataset

There are four major categories of networking attacks. Every attack on a network can be placed into one of these groupings [20].

1) Denial of Service Attack (DoS): is an attack in which the attacker makes some computing or memory resource too busy or too full to handle legitimate requests, or denies\ legitimate users access to a machine.

2) User to Root Attack (U2R): is a class of exploit in which the attacker starts out with access to a normal user account on the system (perhaps gained by sniffing passwords, a dictionary attack, or social engineering) and is able to exploit some vulnerability to gain root access to the system.

3) Remote to Local Attack (R2L): occurs when an attacker who has the ability to send packets to a machine over a network but who does not have an account on that machine exploits some vulnerability to gain local access as a user of that machine.

4) Probing Attack: is an attempt to gather information about a network of computers for the apparent purpose of circumventing its security controls

Two different model Approaches are built for intrusion detection system (Phase-model approach and Level-model approach) that are defined as follows:

1) Phase-Model Approach

Phase model consists of three detection phases. The data is input in the first phase which identifies if this record is a normal record or attack. If the record is identified as an attack then the module inputs this record to the second phase which identifies the class of the coming attack. The second Phase module passes each attack record according to its class type to phase 3 modules. Phase 3 consists of 4 modules one for each class type (DOS, Probe, R2L, U2R). Each module is responsible for identifying the attack type of coming record.

Each Phase was examined with different Decision Tree techniques. The Three Phases are dependent on each other. In other word Phase 2 cannot begin until Phase 1 is finished. This approach has the advantage to flag for suspicious record even if attack type of this record wasn't identified correctly.

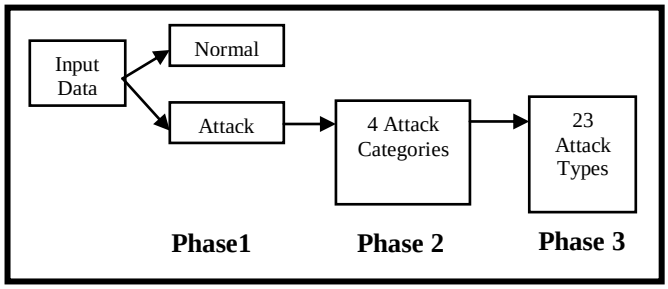


Figure 2. Phase Model Architecture

http://sites.google.com/site/ijcsis/
ISSN 1947-5500

2) Level-Model Approach

Level model consists of 3 independent detection levels. The First Level is to detect normal and Attack profiles. The Second Level is to detect normal records and classify the attacks into four categories independently on the results of the first level. The third Level is to classify each attack type and normal records. Level model approach is to implement each level independent on the other level.

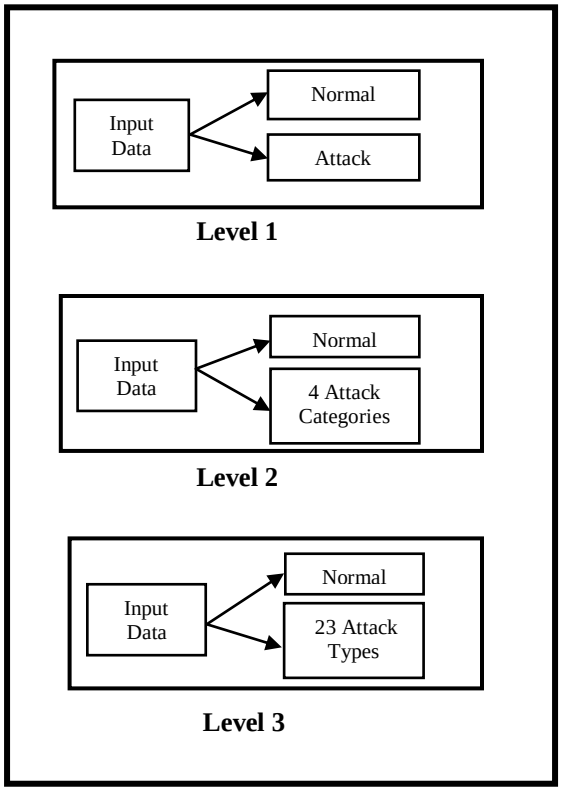


Figure 3. Level Model Architecture

IV. EXPERIMENTS AND RESULTS

A. Data Description

KDDCUP'99 is the mostly widely used data set for the evaluation of these systems. The KDD Cup 1999 uses a version of the data on which the 1998 DARPA Intrusion Detection Evaluation Program was performed. They set up environment to acquire raw TCP/IP dump data for a local area network (LAN) simulating a typical U.S. Air Force LAN.

There are some inherent problems in the KDDCUP'99 data set [21], which is widely used as one of the few publicly available data sets for network-based anomaly detection systems

The data in the experiment is acquired from the NSLKDD dataset which consists of selected records of the complete KDD data set and does not suffer from mentioned shortcomings by removing all the repeated records in the entire KDD train and test set, and kept only one copy of each record [20]. Although, the proposed data set still suffers from some of the problems and may not be a perfect representative of existing real networks, because of the lack of public data sets for network-based IDSs, but still it can be applied as an effective benchmark data set to help researchers compare different intrusion detection methods. The NSL-KDD dataset is available at [22].

We used attacks from the four classes to check the ability of the intrusion detection system to identify attacks from different categories.

The two approaches are examined by two techniques:

1) Test with New Attack: The sample dataset contains 83644 record for training (40000 normal and 43644 for attacks) and 19784 for testing (9647 normal, 6935 for known attacks and 3202 for unknown attacks).

2) Test by Data Partitioning: The sample dataset contain 103427 records is partitioned by 10% (10156 records) for training and 90% (93271 records) for testing.

B. Phase-Module Approach Results

1) Test with New Attack:

Results of Phases model tested with new attacks showed that C5 has a significant detection rate for known and unknown attacks in all phases.

TABLE I. Classification Rate of Phases with New Attacks

Classifier	**Correct Classification Rate**		
	Phase 1	**Phase 2**	**Phase 3**
C5	100 %	**85.34 %**	**99.32%**
CRT	100 %	83.62 %	97.55%
Chaid	100 %	85%	98.73%
Quest	100 %	73.11 %	93.48%

2) Test by Data Patitioning:

Results of data partitioning showed that C5 then CRT & CHAID produced best correct classification rate in second phase which is responsible for classifying coming attack to one of the four classes (DOS, Probe, R2L & U2R). In third phase, C5 showed it has the best classification rate as shown in table II.

http://sites.google.com/site/ijcsis/
ISSN 1947-5500

TABLE II Classification Rate of Phase with Data Partitioning

Classifier	Correct Classification Rate		
	Phase 1	Phase 2	Phase 3
C5	100 %	**99.98 %**	**99.49%**
CRT	100 %	99.97 %	97.02 %
Chaid	100 %	99.79	97.38 %
Quest	100 %	93.74 %	93.25 %

Phase-Model approach has Detection Rate equal to 100 % in both New Attack and Data Partitioning techniques as all attacks in phase 1 are detected correctly.

C. Level-Module Approach Results

1) Test with New Attack:

Testing results showed that C5 produced best correct classification rate for third level and Quest for second level as shown in table III.

TABLE III Classification Rate of Levels with New Attacks

Classifier	Correct Classification Rate		
	Level 1	Level 2	Level 3
C5	100 %	83.82 %	**83.61 %**
CRT	100 %	91.72 %	82.87 %
Chaid	100 %	83.64 %	74.09 %
Quest	100 %	**91.85 %**	77.42 %

TABLE IV Detection Rate of Levels with New Attacks

Classifier	Detection Rate		
	Level 1	Level 2	Level 3
C5	100 %	68.42 %	100 %
CRT	100 %	100 %	100 %
Chaid	100 %	68.41 %	93.42 %
Quest	100 %	100 %	100 %

2) Test by Data Patitioning:

Results of data partitioning showed that second level are easy to be correctly classified by many decision trees classifiers either C5, CRT or CHAID. In third phase, C5 showed it has the best classification rate as shown in table V.

TABLE V Classification Rate of Levels with Data Partitioning

Classifier	Correct Classification Rate		
	Level 1	Level 2	Level 3
C5	100 %	**99.96 %**	**99.73 %**
CRT	100 %	99.89 %	90.22 %
Chaid	100 %	99.88 %	87.92 %
Quest	100 %	97.17 %	88.28 %

TABLE VI Detection Rate of Levels with Data Partitioning

Classifier	Detection Rate		
	Level 1	Level 2	Level 3
C5	100 %	99.92 %	100 %
CRT	100 %	100 %	100%
Chaid	100 %	99.92 %	96.52 %
Quest	100 %	100 %	100 %

V. DISCUSSION

We defined two different Approaches. The first approach is the phase model approach which consists of three sequential detection levels. Phase 1 is able to detect Normal and Attack behavior. Phase 2 is to classify the attacks detected from phase 1 into 4 Attack categories (DOS, Probe, R2L, U2R). Phase 3 is to classify each attack type in each category.

The second approach is the level model approach which consists of 3 separated detection level. Level1 is to detect normal and Attack profiles. Level2 is to detect normal records and classify the attacks into four categories. Level3 is to classify each attack type and normal records.

TABLE VII Comparison between Phase and Level approaches

	Phase Approach	Level Approach
Training Time	less training time	High training time
Detection Rate	Higher detection Rate for New Attacks	Lower detection rate for New Attacks
False Alarm Rate (FAR)	Lower FAR as Attacks are detected in the first phase	Higher FAR as Attacks Types and Categories a are detected in parallel with the normal records
Errors Propagation	May propagate errors	Does not propagate errors
Classification Rate	Higher Classification Rate in New Attacks and Data Partitioning Techniques	Lower classification Rate in New Attacks technique.

As shown in table VII, Phase model take less training time and even decrease in each phase where we use the whole dataset for training phase 1 then in phase 2 we use only the attacks for training excluding the normal records. While in Level model, it takes high training time as the whole data is entered in the training of each level.

Phase model has higher detection Rate for New Attacks which never been seen before but lower detection rate for New Attacks in level model.

Attacks are detected in the first phase then are sent for further classification to the next phase without Normal records

http://sites.google.com/site/ijcsis/
ISSN 1947-5500

but in Level model, Attacks Types and Categories are detected in parallel with the normal records which may increase the false alarm rate.

Phase model May propagate errors as each phase is dependent on the previous one. But level model does not propagate errors as each level is separated and has independent results.

Phase model has Higher Classification Rate in New Attacks and Data Partitioning Techniques than Level model which has Lower classification Rate in New Attacks technique.

VI. CONCLUSION AND FUTURE WORK

In this paper we compared the results of 2 different approaches of intrusion detection system (Phase and Level Approach). Phase Approach consists of three detection phases. The data is input in the first phase which identifies if this record is a normal record or attack. If the record is identified as an attack then the module inputs this record to the second phase which identifies the class of the coming attack. The second phase module passes each attack record according to its class type to phase 3 modules. Phase 3 consists of 4 modules one for each class type (DOS, Probe, R2L, U2R). Each module is responsible for identifying the attack type of coming record. While the Level approach consists of 3 independent detection levels. The First Level is to detect normal and Attack profiles. The Second Level is to detect normal records and classify the attacks into four categories independently on the results of the first level. The third Level is to classify each attack type and normal records.

We examined each model approach using different decision trees modules (C5, CRT, QUEST and CHAID). Each module is implemented by applying 2 techniques (New Attacks and Data Partitioning Techniques) .First, New Attacks Technique is to add new attacks in testing. Second, Data Partitioning Technique is to divide the dataset into 10 %for training and 90% for testing.
New Attacks technique is more realistic than Data Partitioning technique as in real life we are exposed to new attacks every second which we can't expect.

The results show that C5 decision tree has the most significant detection rate for both phase and level approaches. CRT & CHAID have promising results in Data Partitioning technique for both phase and level approaches.
Quest has high classification rate when adding new attacks in the second level.

The experimental results showed that Phase Model approach has Higher Classification Rate in New Attacks and Data Partitioning Techniques than Level Model approach. Therefore, the phase approach is more realistic than Level approach as in real life we are exposed every second to new attacks that we don't expect.

The Future work will be directed towards finding ways to prevent propagating errors in phase model. Also using other Machine learning techniques in our experiments for detecting more types of intrusions.

REFERENCES

[1] T. Shon and J. Moon, "A hybrid machine learning approach to network anomaly detection", Information Sciences, vol.177, pp. 3799-3821, 2007.

[2] Mostafa Salama, Heba Eid, Ashraf Darwish Aboul Ella Hassanien Hybrid Intelligent Intrusion Detection Scheme. 15th Online World Conference on Soft Computing in Industrial Applications, 15th to 27th November 2010, Springer in "Advances In Intelligent and Soft Computing,2010

[3] Z.S. Pan, S.C. Chen, G.B Hu and D.Q. Zhang, "Hybrid Neural Network and C4.5 for Misuse Detection," In Machine Learning and Cybernetics, pp. 2463-2467. Xi'an, 2003.

[4] J.P. Anderson, Computer Security Threat Monitoring and Surveillance, http://csrc.nist.gov/publications/history/ande80.pdf, 2010.

[5] Vyas Sekar, Ravishankar Krishnaswamy, Anupam Gupta, Michael K. Reiter, "Network-Wide Deployment of Intrusion Detection and Prevention Systems", 2010

[6] Naelah okasha, Abd El Fatah Hegazy, Sherif M. Badr, 2010. "Towards Ontology-Based Adaptive Multilevel Model for Intrusion Detection and Prevention System (AMIDPS)", Egyptian science journal (ESC), Vol. 34, No. 5, September 2010.

[7] R. Bace and P. Mell, Intrusion Detection Systems, Computer Security Division, Information Technology Laboratory, Nat'l Inst. of Standards and Technology, 2001.

[8] Kapil Kumar Gupta, Baikunth Nath, and Ramamohanarao Kotagiri "Layered Approach Using Conditional Random Fields for Intrusion Detection" IEEE Transactions on dependable and secure Computing, vol. 5, no. 4, october-december 2008.

[9] Sahar Selim, Mohamed Hashem and Taymoor M. Nazmy, "Hybrid Multi-level Intrusion Detection System", International Journal of Computer Science and Information Security (IJCSIS), pp. 23-29, Vol. 9, No. 5, May 2011

[10] Asmaa Shaker Ashoor, Prof. Sharad Gore,"Importance of Intrusion Detection System (IDS)", International Journal of Scientific & Engineering Research (IJSER), Volume 2, Issue 1, January-2011.

[11] N.B. Amor, S. Benferhat, and Z. Elouedi, "Naive Bayes vs. Decision Trees in Intrusion Detection Systems," Proc. ACM Symp. Applied Computing (SAC '04), pp. 420-424, 2004.

[12] T. M. Mitchell. Machine Learning. McGraw Hill, 1997 .

[13] Quinlan JR. "C4.5: programs for machine learning," Log Altos,CA: Morgan Kaufmann; 1993.

[14] SPSS. Clementine 12.0 modeling nodes. Chicago: SPSS; 2007 .

[15] L. Brieman, J. Friedman, R. Olshen and C. Stone, "Classification of Regression Trees," Wadsworth Inc., 1984.

[16] J.A. Michael and S.L. Gordon, "Data mining technique for marketing, sales and customer support," Wiley, New York, 1997.

[17] W. Y. Loh and Y. S. Shih, "Split selection methods for classification trees, "Statistica Sinica 7, pp. 815–840, 1997.

[18] Sahar Selim, M. Hashem and Taymoor M. Nazmy, "Intrusion Detection using Multi-Stage Neural Network, " International Journal of Computer Science and Information Security, Vol. 8, No. 4, 2010.

[19] S.T. Sarasamma, Q.A. Zhu, and J. Huff, "Hierarchal Kohonenen Net for Anomaly Detection in Network Security," IEEE Transactions on Systems, Man, and Cybernetics-Part B: Cybernetics, 35(2), 2005, pp.302-312.

http://sites.google.com/site/ijcsis/
ISSN 1947-5500

[20] M. Tavallaee, E. Bagheri, W. Lu, and A. Ghorbani, "A Detailed Analysis of the KDD CUP 99 Data Set," Submitted to Second IEEE Symposium on Computational Intelligence for Security and Defense Applications (CISDA), 2009.

[21] KDD Cup 1999. Available on: http://kdd.ics.uci.edu/databases/kddcup 99/kddcup99.html, October 2007

[22] "NSL-KDD data set for network-based intrusion detection systems", Available on: http://nsl.cs.unb.ca/NSL-KDD/, March 2009.

[23] Mohammad Sazzadul Hoque, Md. Abdul Mukit and Md. Abu Naser Bikas," An Implementation of Intrusion Detection System using Genetic Algorithm ", International Journal of Network Security & Its Applications (IJNSA), Vol.4, No.2, March 2012.

AUTHORS PROFILE

Heba Ezzat Ibrahim Bachelor of Computer Science. Currently working for master degree in Arab Academy for Science and Technology & Maritime Transport.

Sherif M. Badr PHD degree in Computer Engineering in Military Technical College. Fields of interest are intrusion detection, computer and networks security

Mohamed A. Shaheen Associate Professor in College of Computing and Information Technology in Arab Academy for Science and Technology & Maritime Transport

Survey on Using GIS in Evacuation Planning Process

Sara Shaker Abed El-Hamied
Information Systems Department
Computer and Information Sciences Faculty
Mansoura, Egypt
sara.shaker2008@yahoo.com

Ahmed Abou El-Fotouh Saleh
Information Systems Department
Computer and Information Sciences Faculty
Mansoura, Egypt
elfetouh@gmail.com

Aziza Asem
Information Systems Department
Computer and Information Sciences Faculty
Mansoura, Egypt
aziza_asem2@hotmail.com

***Abstract*- Natural crises form a big threat on environment; these crises mean the loss of enterprises and individuals, and therefore losses in the sum total of community development. Management to these crises is required through a crisis management plan to control the crises before, during, and after the event. One of the most needed things to consider during preparing the crisis management plan is preparing the evacuation plan in order to transfer people from the incident place to a safe place; this must be done quickly and carefully.**

Because of the geographic nature of the evacuation process, Geographical Information System (GIS) has been used widely and effectively for over 20 years in the field of crisis management in general and in evacuation planning in particular. This paper provides an overview about evacuation process and the basic concepts of GIS systems. The paper also demonstrates the importance of evacuation planning and how GIS systems used in other studies to assists in evacuation process

Keywords-Crisis Management; Evacuation Planning; Geographical Information System (GIS).

I. INTRODUCTION

Crises always threaten society; they happen suddenly and cause significant losses. Crisis management is the process of controlling crises before, during, and after the event. One of crisis management activities is Evacuation planning, which means the transfer of people from an unsafe place to another safe place.

Within crises management, emergency management applies geo-information technologies in the crisis management process and Geographical Information Systems (GIS) have been used for over 20 years. Examples for GIS utilization in natural and man-made disasters are to support flood mapping, hurricane prediction, and environmental clean-ups after industrial accidents.

In most crises situations, GIS operators receive their orders via staff members who are asked by the decision makers to inquire about maps. The GIS specialists usually react to mapping and spatial analysis requests from decision makers, e.g. after the World Trade Center attack GIS specialists, supported by company consultants, were operating Geographic Information Systems and producing maps on demand and, after Hurricane Katrina, GIS experts from Louisiana State University provided support to evacuation and relief efforts. In larger communities, state and federal agencies GIS operations have become an integral part of Emergency Operation Centers (EOCs), but in some instances, e.g. smaller and/or rural communities, special GIS operators might not be available or are not part of the Emergency Operation Center staff. Even in New Orleans, a major metropolitan area, GIS use was hindered during Hurricane Katrina because the mapping requests overwhelmed the EOC capabilities and outdated computers caused frustration [1].

II. EVACUATION PLANNING

Crises mean the loss of enterprises and individuals, and therefore losses in the sum total of community development, crises have several types such as nature, industrial, technological, etc. Crisis management is the process to control the crisis by developing plans to reduce the risk of a crisis occurring and to deal with any crises that arise, and the implementation of these plans so as to minimize the impact of crises and assist the organization to recover from them.

Evacuation is one of the crises management activities which is an operation where by all or part of a particular population is temporary relocated, whether spontaneously or in an organized manner, from a sector that has been struck by a disaster or is about to be struck by a disaster, to a place considered not dangerous for its health or safety [11]. Evacuation can be carried for several reasons such as volcanoes, floods, hurricanes, earthquakes, military attacks, industrial accidents, traffic accidents, fire, nuclear accidents … etc.

Evacuation Plan is a supporting document that is used to identify and organize the various responses aimed at evacuating persons exposed to a threat from an evacuation sector to a reception sector, while ensuring them a minimum of essential services during an emergency. Proper planning will use multiple exits, contra-flow lanes, and special technologies to ensure full, fast and complete evacuation and should consider personal situations which may affect an individual's ability to evacuate. These plans may also include alarm signals that use both aural and visual alerts.

http://sites.google.com/site/ijcsis/
ISSN 1947-5500

All countries should have a written evacuation plan in order to facilitate a safe and efficient evacuation or relocation for their people and the plan must be updated regularly. Country Directors must communicate in writing what evacuation assistance will provide for each member of the staff and their families in the event of a crisis [2].

Evacuation planning has been studied from different perspectives such as evacuee's behaviors, traffic control, safe area selection, and route finding to safe areas [3], with any perspective an evacuation plan should involve four phases [2]:

1. **Pre-Planning:** During this phase, operations are normal with periodic update. The country office must ensure continual monitoring of the safety and security situation, especially in high risk areas.
2. **Alert:** Mounting tension may lead the country director to issue a recommendation to limit operations, increase security measure, and review the evacuation plan.
3. **Curtailment of operations/relocation:** The situation has deteriorated to a level unsafe for normal operations and may require rapid evacuation.
4. **Evacuation:** The planned evacuation process become in effect and all threatened people must be transferred to safe areas.

III. GEOGRAPHICAL INFORMATION SYSTEMS

As sun rises and sets, people everywhere ask questions about locations on earth like: Where can I find the shop? Where is the nearest library? How can I go to the restaurant? Which site is the best site for the building? All of these questions and more can be asked by a Geographical Information System (GIS). GIS involves handling the issues arising from working with geographic information, also it examines the effect of GIS on people and society, and the effect of society on GIS.

There have been so many attempts to define GIS that make it difficult to select one definitive definition [4] because the definition will depend on the one giving it and his point of view. For example, *Rhind D.W. (1989)* defines it generally as "a computer system that can hold and use data describing places on the Earth's surface" [5]. Other definitions explain what a GIS can do. For example, *Burrough P.A. (1986)* define a GIS as "a set of tools for collecting, storing, retrieving at will, transforming, and displaying spatial data from the real world for a particular set of purposes" [6], the US Government defines it as "a system of computer software and procedures designed to support the capture, management, manipulation, analysis, and display of spatially referenced data for solving complex planning and management problems." [12], and the *Department of the Environment (1987)* say that a GIS is "a system for capturing, storing, checking, integrating, manipulating, analyzing and displaying data which are spatially referenced to the earth" [7].

Simply, GIS system considers three main components: hardware, software, and spatially referenced data. In particular, a working GIS needs to integrate five components:

1. *Hardware*

   GIS hardware includes a computer with high capabilities on which a GIS operates, a monitor on which results displays, a printer to display results as reports, other GIS hardware also includes GPS instrument to collect coordinates, and a digitizer.

2. *Software*

   Key software components are:

   - System software (e.g. operating system).
   - a database management system (DBMS)
   - tools for the input and manipulation of geographic information
   - tools that support geographic query, analysis, and visualization
   - a graphical user interface (GUI) for easy access to tools
   - Drawing software.

3. *People*

   GIS people can be divided into two main categories:
   - People who develop the GIS and define its tasks such as database administrators, application specialists, systems analysts, and programmers. They are responsible for maintenance of the geographic database and provide technical support.
   - General users who are using the GIS in their daily business.

4. *Data*

   May be the most important part of a GIS, a GIS integrates spatial data with other attribute data to answer unique spatial queries provided by users.
   - Spatial data: data can be referenced to a location on earth. For example, country, road, river, etc.
   - Attribute data: also called aspatial data, data linked to spatial data describe those data. For example, country name, road length, river width, etc.

5. *Methods*

   A well defined consistent rules that the GIS needs to achieve its goals includes how the data will be retrieved, input into the system, stored, managed, transformed, analyzed, and finally presented in a final output.

From information systems point of view, GIS is like any other information system that considers three main phases: input, processing, and output. GIS helps in answering questions and solving problems by looking at the data in a way that is quickly understood and easily shared. GIS technology

http://sites.google.com/site/ijcsis/
ISSN 1947-5500

can be integrated into any enterprise information system framework [13].

What distinguish GIS system from any other information system is that it combines a powerful visualization environment with a strong analytical and modeling framework, which makes GIS attractive to most people in the whole world. For example, when a rainfall occur it is important to know where it is located. By using a spatial reference system such as latitude, longitude, or elevation we can know where the rainfall, and by comparing the results with the landscape one can predicates if there are someplace that likely to be subjected to dry up.

Designing a GIS system is a sequenced process begins with the data collection phase in which data; spatial and attribute data, from various data sources is collected, these spatial data must be geo-referenced to their spatial location on the earth. Then these data are digitized to convert it from the analog format into the digital one by a trace method. After that, attribute data is combined with the spatial data into a data format so that it can be manipulated by the system to provide answers that help decision makers take their decisions. This process is utilized in Figure 1.

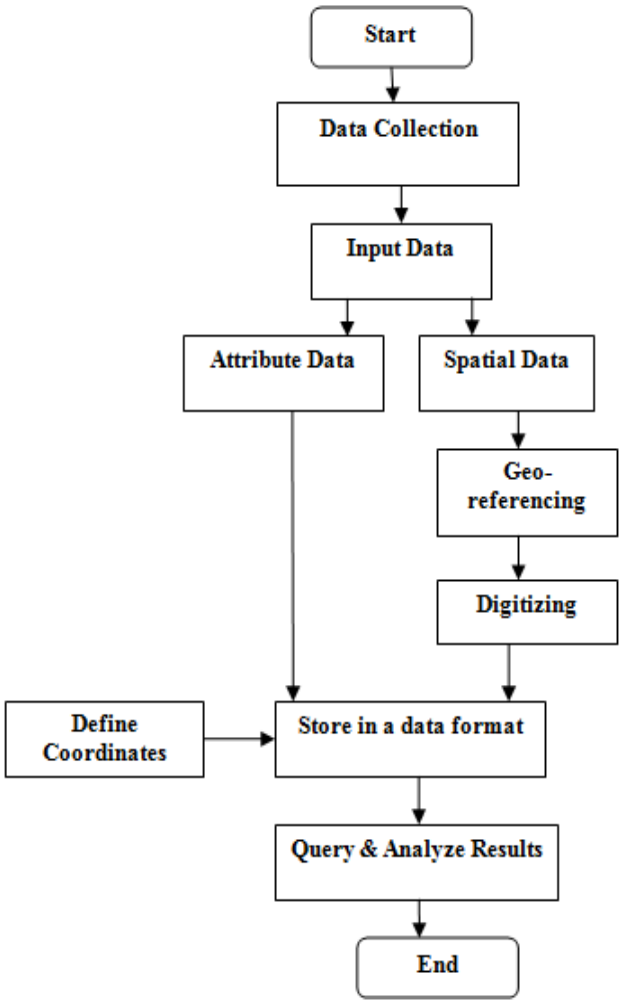


Figure 1. GIS system's design process.

IV. USING GIS IN EVACUATION PLANNING

Evacuation is a process in which threatened people are transferred from the incident place to a safer place to protect their lives. It is a very complex process, besides needing to be accurate and carful; it must be done very quickly. GIS plays an important role in emergency management in general and in evacuation planning in particular.

In 2001 GIS was prominently used in the rescue, relief and recovery process after the World Trade Center attack. Although New York City's Emergency Operation Center and GIS infrastructure was destroyed, city officials were able to set up a backup facility and use GIS to produce maps for emergency response purposes by the evening of 9/11[1]

Yang Bo et al (2009) assure that emergency evacuation is an important measure for preventing and reducing injuries and death during large scale emergency. They assumed that the efficiency of evacuation is based on (1) Understanding of the situation, and (2) Good analysis and judgment of information.

So they proposed a multi-agent framework and a GIS system that:

1. Simulate individual movement by a modified Particle Swarm Optimization (LWDPSO) which considers each individual as a particle and as one particle found an exit all other particles should consider all other exits as an exit and choose the nearest one, and
2. Modeling the evacuation environment by a GIS platform, in which each individual takes an average space 0.4m × 0.4m when it is very crowded, and then build a potential map for evacuation environment which describes the distance between an individual and an exit [8].

Michel Pidd et al (1993) developed a Configurable Evacuation Management and Planning System (CEMPS) to be used for evacuation from man-made disasters. They found GIS as an efficiency technology that can examine static aspects of an evacuation plan such as determining evacuation zone and evacuation routes, but they assume that it can't consider dynamic aspects such as How long vehicles take to come? How long will it take to evacuate the population? So they use micro-simulation method in order to simulate the movement of people from the evacuation zone to the vehicles and use ArcInfo to determine the evacuation zone and evacuation routes [9].

Mohammad Saadatseresht et al (2008) introduce that the distribution of population into safe areas during evacuation process is a vital problem that affect the efficiency of the evacuation plan. They propose a three step approach in order to determine the distribution of evacuees into the safe areas, in which step1: is to select the safe areas, step2: is to find optimal path between each building block and the candidates safe areas, and step3: is to select optimal safe areas for each building block, optimal safe area should be the closest to the building block and should have enough space for evacuees. To

http://sites.google.com/site/ijcsis/
ISSN 1947-5500

achieve the third step two objective functions were defined, and then the spatial MOP was solved using the NSGA-II algorithm in a GIS environment [3].

Bo Huang and Xiaohong Pan (2006) introduce an Incident Response Management Tool (IRMT) in order to reduce response time in incident management. The IRMT consists of:

1. GIS system: provide user interface, process network data, find shortest path, and visualize the result.
2. Traffic simulation engine: simulate incidents, gather link travel time at regular intervals, and send this time to the GIS system.

Optimization engine: to minimize the overall travel time of all response units [10].

ACKNOWLEDGMENT

First of all, Praise and thank Allah that his grace is righteous, and then I thank my family and my friends for supporting me.

REFERENCES

[1] Sven Fuhrmann, Alan MacEachren, and Guoray Cai, "Geoinformation Technologies to Support Collaborative Emergency Management", chapter20, Springer, 2008.

[2] CARE International, "Evacuation", Security & Safety Manual, chapter 6.

[3] Mohammad Saadatseresht, Ali Mansourian, and Mohammad Taleai, "Evacuation Planning Using Multiobjective Evolutionary Optimization Approach", European Journal of Operation Research, 2008.

[4] Ian Heywood, Sarah Cornelius, Steve Carver, "An Introduction to Geographical Information Systems", 3rd edition, Pearson, 2006.

[5] Rhind D W, "Why GIS?" pp. 9-28, ARC News, 1989.

[6] Burrough P A, "Principles of Geographical Information Systems for Land Resources Assessment", Clarendon press, 1986.

[7] Department of the Environment, "Handling Geographic Information", HMSO, London, 1987.

[8] Yang Bo, Wu Yong-gang, and Wang Cheng, "A multi-agent and GIS based simulation for emergency evacuation in park and public square", IEEE, 2009.

[9] Michael Pidd, F.N. de Silva, Richard W.Eglese, "CEMPS: A Configurable Evacuation Management System- a Progressive Report", Winter Simulation Conference, 1993.

[10] Bo Huang, and Xiaohong Pan, "GIS coupled with traffic simulation and optimization for incident response", ScienceDirect, 2006

[11] The City of Brampton Emergency Evacuation plan 2011, http://www.brampton.ca/EN/RESIDENTS/Pages/Welcome.aspx

[12] http://geology.wlu.edu/harbor/geol260/lecture_notes/notes_intro1.html

[13] http://gis.com/content/what-gis

http://sites.google.com/site/ijcsis/
ISSN 1947-5500

Classification and Importance of Intrusion Detection System

K Rajasekaran
Research Scholar,
Research & Development Centre,
Bharathiar University, Coimbatore,
India
krs.salem@gmail.com

Dr. K Nirmala
Associate Professor in Computer Science,
Quiad-E-Millath Govt. College for Women, Chennai,
India.
nimimca@gmail.com

Abstract:- An intrusion detection system (IDS) is a device or software application that monitors network or system activities for malicious activities or policy violations and produces reports to a Management Station. Some systems may attempt to stop an intrusion attempt but this is neither required nor expected of a monitoring system. Due to a growing number of intrusion events and also because the Internet and local networks have become so ubiquitous, organizations are increasingly implementing various systems that monitor IT security breaches. This includes an overview of the classification of intrusion detection systems and introduces the reader to some fundamental concepts of IDS methodology: audit trail analysis and on-the-fly processing as well as anomaly detection and signature detection approaches. This research paper discusses the primary intrusion detection techniques and the classification of intrusion Detection system.

Keywords: Intrusion Detection, signature, anomaly, specification, classification

I. INTRODUCTION

The main aim of intrusion detection is to monitor network assets to detect anomalous behaviour and misuse in network. Intrusion Detection has been around for nearly twenty years but only recently has it seen a dramatic rise in popularity and incorporation into the overall information security Infrastructure. Beginning in the year 1980's James Anderson's seminal paper, was written for a government organization, introduced the notion that audit trails contained vital information that could be valuable in tracking misuse and understanding of user behaviour. With the release of Anderson's paper, the concept of "detecting" misuse and specific user events emerged. His work was the start of host-based intrusion detection technique and IDS in general.

Commercial development of intrusion detection technologies began in the year of 1990s. Haystack Labs was the first commercial vendor of IDS tools, with its Stalker line of host-based products. SAIC was also developing a form of host-based intrusion detection; this system is called Computer Misuse Detection system (CMDS).

Simultaneously, the Air Force's Crypto logic Support Center developed the Automated Security Measurement System (ASIM) to monitor network traffic on the US Air Force's network. ASIM also made considerable progress in overcoming scalability and portability issues that previously plagued NID products. Additionally, ASIM was the first solution to incorporate both a hardware and software solution to network intrusion detection technique. ASIM is currently in use and managed by the Air Force's Computer emergency Response Team (AFCERT) at locations all over the world. As often happened, the development group on the ASIM project formed a commercial company in 1994, the Wheel Group. Their product, Net Ranger, was the first commercially viable network intrusion detection device management system.

The intrusion detection market began to gain in popularity and truly generate revenues around 1997. In that year, the security market leader, ISS, developed a network intrusion detection system called Real Secure. A year later, Cisco recognized the importance of network intrusion detection and purchased the Wheel Group, attaining a security solution they could provide to their customers. Similarly, the first visible host-based intrusion detection company, Centrex Corporation, emerged as a result of a merger of the development staff from Haystack Labs and the departure of the CMDS team from SAIC. From there, the commercial IDS world expanded its market-base and a roller coaster ride of start-up companies, mergers, and acquisitions ensued.

http://sites.google.com/site/ijcsis/
ISSN 1947-5500

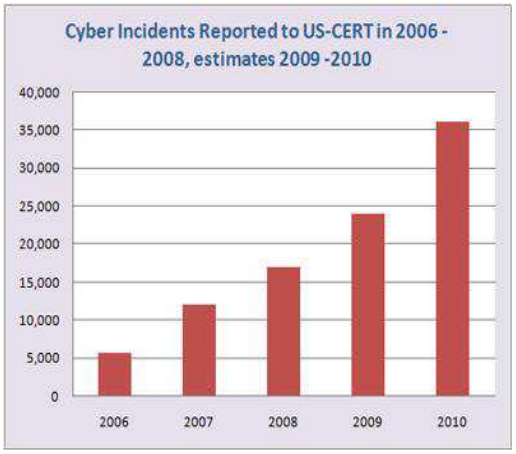


Figure 1: Number of incidents reported

The above chart from US-CERT shows how the cyber incidents rose in current internet network environment; this gives requirement of IDS deployment in network security system.

II. INTRUSION DETECTION SYSTEM TECHNIQUES

Intuitively, intrusions in an information system are the activities that violate the security policy of the system, and intrusion detection is the process used to identify intrusions. Intrusion detection has been studied for approximately 20 years. It is based on the beliefs that an intruder's behavior will be noticeably different from that of a legitimate user and that many unauthorized actions will be detectable. Intrusion detection system is classified into three categories. The different types of intrusion Detection techniques are listed below.

A. Signature based detection systems,

B. Anomaly based intrusion detection systems

C. Specification based detection systems.

A. Signature based detection systems

Signature based detection system (also called misuse based), this type of detection is very effective against known attacks, and it depends on the receiving of regular updates of patterns and will be unable to detect unknown previous threats or new releases.

One big challenge of signature-based IDS is that every signature requires an entry in the database, and so a complete database might contain hundreds or even thousands of entries. Each packet is to be compared with all the entries in the database. This can be very resource- consuming and doing so will slow down the throughput and making the IDS vulnerable to DoS attacks. Some of the IDS evasion tools use this vulnerability and flood the signature signature-based IDS systems with too many packets to the point that the IDS cannot keep up with the traffic, thus making the

IDS time out and drop packets and as a result, possibly miss attacks. Further, this type of IDS is still vulnerable against unknown attacks as it relies on the signatures currently in the database to detect attacks.

B. Anomaly based detection system

This type of detection depends on the classification of the network to the normal and anomalous, as this classification is based on rules or heuristics rather than patterns or signatures and the implementation of this system we first need to know the normal behaviour of the network. Anomaly based detection system unlike the misuse based detection system because it can detect previous unknown threats, but the false positive to rise more probably.

The signature of a new attack is not known before it is detected and carefully analyzed. So it is difficult to draw conclusions based on a small number of packets. In this case, anomaly-based systems detect abnormal behaviors and generate alarms based on the abnormal patterns in network traffic or application behaviors. Typical anomalous behaviors that may be captured include

1) Misuse of network protocols such as overlapped IP fragments and running a standard protocol on a stealthy port;

2) Uncharacteristic traffic patterns, such as more UDP packets compared to TCP ones,

3) Suspicious patterns in application payload.

The big challenges of anomaly based detection systems are defining what a normal network behavior is, deciding the threshold to trigger the alarm, and preventing false alarms. The users of the network are normally human, and people are hard to predict. If the normal model is not defined carefully, there will be lots of false alarms and the detection system will suffer from degraded performance.

C. Specification based detection system

This type of detection systems is responsible for monitoring the processes and matching the actual data with the program and in case of any abnormal behaviour will be issued an alert and must be maintained and updated whenever a change was made on the surveillance programs in order to be able to detect the previous attacks the unknown and the number of false positives what can be less than the anomaly detection system approach.

http://sites.google.com/site/ijcsis/
ISSN 1947-5500

III. CLASSIFICATION OF INTRUSION DETECTION SYSTEM

When considering the area being the source of data used for intrusion detection, another classification of intrusion detection systems can be used in terms of the type of the protected system. There is a family of IDS tools that use information derived from a single host (system) — host based IDS (HIDS) and those IDSs that exploit information obtained from a whole segment of a local network (network based IDS, i.e. NIDS) and the combined Hybrid based Intrusion Detection system. Intrusion detection system is mainly classified into three types. The classification of intrusion Detection systems are listed below:

a. Host based IDS

b. Network based IDS

c. Hybrid based IDS

A. *Host based IDS (HIDS)*

This type is placed on one device such as server or workstation, where the data is analyzed locally to the machine and are collecting this data from different sources. HIDS can use both anomaly and misuse detection system.

A Host Intrusion Detection Systems (HIDS) and software application (agents) installed on workstations which are to be monitored. The agents monitor the operating system and write data to log files and/or trigger alarms. A host Intrusion detection systems (HIDS) can only monitors the individual workstations on which the agents are installed and it cannot monitor the entire network. Host based IDS systems are used to monitor any intrusion attempts on critical servers.

The drawbacks of Host Intrusion Detection Systems (HIDS) are

• Difficult to analyse the intrusion attempts on multiple computers.

• Host Intrusion Detection Systems (HIDS) can be very difficult to maintain in large networks with different operating systems and configurations

• Host Intrusion Detection Systems (HIDS) can be disabled by attackers after the system is compromised.

Systems that monitor incoming connection attempts (RealSecure Agent, PortSentry). These examine host-based incoming and outgoing network connections. These are particularly related to the unauthorized connection attempts to TCP or UDP ports and can also detect incoming portscans.

Systems that examine network traffic (packets) that attempts to access the host. These systems protect the host by intercepting suspicious packets and looking for aberrant payloads (packet inspection).

Systems that monitor login activity onto the networking layer of their protected host (HostSentry). Their role is to monitor log-in and log-out attempts, looking for unusual activity on a system occurring at unexpected times, particular network locations or detecting multiple login attempts (particularly failed ones).

The HIDS that look only at their host traffic can easily detect local-to-local attacks or local-to-root attacks, since they have a clear concept of locally available information, for example they can exploit user IDS. Also, anomaly detection tools feature a better coverage of internal problems since their detection ability is based on the normal behavior patterns of the user.

The HIDS reside on a particular computer and provide protection for a specific computer system. They are not only equipped with system monitoring facilities but also include other modules of a typical IDS.

HIDS products such as Snort, Dragon Squire, Emerald eXpert-BSM, NFR HID, Intruder Alert all perform this type of monitoring.

B. *Network based IDS (NIDS)*

Network Intrusion Detection Systems (NIDS) usually consists of a network appliance (or sensor) with a Network Interface Card (NIC) operating in promiscuous mode and a separate management interface. The IDS is placed along a network segment or boundary and monitors all traffic on that segment.

NIDS are deployed on strategic point in network infrastructure. The NIDS can capture and analyze data to detect known attacks by comparing patterns or signatures of the database or detection of illegal activities by scanning traffic for anomalous activity. NIDS are also referred as “packet-sniffers”, Because it captures the packets passing through the of communication mediums.

The network-based type of IDS (NIDS) produces data about local network usage. The NIDS reassemble and analyze all network packets that reach the network interface card operating in promiscuous mode. They do not only deal with packets going to a specific host – since all the machines in a network segment benefit from the protection of the NIDS. Network-based IDS can also be installed on active network elements, for example on routers.

Since intrusion detection (for example flood-type attack) employs statistical data on the network load, a certain type of dedicated NIDS can be separately

http://sites.google.com/site/ijcsis/
ISSN 1947-5500

distinguished, for example, those that monitor the traffic (Novell Analyzer, Microsoft Network Monitor). These capture all packets that they see on the network segment without analyzing them and just focusing on creating network traffic statistics.

Typical network-based intrusion systems are: Cisco Secure IDS (formerly NetRanger), Hogwash, Dragon, E-Trust IDS.

C. *Hybrid based IDS*

The management and alerting from both network and host based intrusion detection devices, and provide the logical complement to NID and HID - central intrusion detection management. Both Network and Host based IDS have their own unique advantages and disadvantages. Network based IDS are easier to deploy and are less expensive to purchase and maintain. However, their performance depends on known security exploits and signatures. If a new exploit is used that the IDS is unaware of, the system could easily fail to detect the attack. A host based IDS is only as good as the security administrator who maintains and monitors it. Becoming skilled at, maintaining and monitoring this software can be a daunting task. Therefore, the best approach is to use a combination of the best features of Network based and Host based IDS to improve resistance to attacks and to provide greater flexibility. This approach is commonly referred to as Hybrid IDS.

IV. CONCLUSION

Intrusion detection continues to be an active research field .An intrusion detection system is a part of the defensive operations that complements the defences such as firewalls, UTM etc. The intrusion detection system basically detects attack signs and then alerts. According to the detection methodology, intrusion detection systems are typically categorized as misuse detection and anomaly detection systems. The deployment perspective, they are be classified in network based or host based IDS. In current intrusion detection systems where information is collected from both network and host resources. Moreover, reconstructing attack scenarios from intrusion alerts and integration of IDSs will improve both the usability and the performance of IDSs. Many researchers and practitioners are actively addressing these problems In terms of performance; an intrusion detection system becomes more accurate as it detects more attacks and raises fewer false positive alarms.

REFERENCE

[1] Anderson D, Lunt TF, Javitz H, Tamaru A, Valdes A. Detecting unusual program behaviour using the statistical component of the next-generation intrusion detection expert system (NIDES). Menlo Park, CA, USA: Computer Science Laboratory, SRI International; 1995. SRIO-CSL-95-06.

[2] Ghosh, A.K., Wanken, J., & Charron, F. Detecting anomalous and unknown intrusions against programs. In K. Keus (Ed), *Proceedings of the 14th* annual computer security applications conference , 1998, (pp. 259--267). IEEE Computer Society, Los Alamitos,CA.

[3] G. Macia Fernandez and E. Vazquez, "Anomaly-based Network intrusion detection: Techniques, systems and Challenges", Computers & Security, Vol. 28, No. 1-2, pp. 18-28, February-March 2009.

[4] Harley Kozushko, "Intrusion Detection: Host-Based and Network-Based Intrusion Detection Systems", on September 11, 2003.

[5] Paul Innella Tetrad, "The Evolution of Intrusion Detection Systems", Digital Integrity,LLC on November 16, 2001.

[6] Rasha G. Mohammed Helali, "Data Mining Based Network Intrusion Detection System: A Survey", In Novel Algorithms and Techniques in Telecommunications and Networking, pp. 501-505, 2010.

[7] Pakkurthi Srinivasu, P.S. Avadhani, Vishal Korimilli, Prudhvi Ravipati, "Approaches and Data Processing Techniques for Intrusion Detection Systems", Vol. 9, No. 12, pp. 181-186, 2009.

[8] Sekar R., Gupta A., Frullo J., Shanbhag T., Tiwari A., Yang H., et al. Specification-based anomaly detection: a new approach for detecting network intrusions. In: Proceedings of the Ninth ACM Conference on Computer and Communications Security; 2002. p. 265–74.

Authors Profile:

K.Rajasekaran received his B.Sc. Degree in computer Science from Vysya College, Salem,India and M.C.A. Degree form K.S.R. College of Technology, Tiruchengode, India. He also received his M.Phil Degree in computer science from Periyar University. He is now doing his Ph.D. in computer science at Research and Development Centre, Bhrathiar University, Coimbatore, India. His field of interest is Networks, Data Mining and computer Architecture.

Dr. K.Nirmala received her Ph.D. Degree in Computer Science from NITTTR, Taramani, University of Madras, Chennai, India. She has fifteen years of teaching experience in the field of Computer Science at college level education. Since 1997 she has been working in various levels in the department of higher education, Tamilnadu, India. She is now working as Associate Professor of Computer Science, Quaid-E-millath Govt. College for Women, Chennai, India. Her field of interest is Data mining, Networks and Operating System. She has presented and published many technical papers at various national and international conferences and journals.

http://sites.google.com/site/ijcsis/
ISSN 1947-5500

Elimination of Weak Elliptic Curve Using Order of Points

Nishant Sinha[#1], Aakash Bansal[*2]

[#] *School of IT*

CDAC Noida, India

[1]sinha22nishant@gmail.com

[*]*School of IT*

CDAC Noida, India

[2]aakashbansal.cdac@gmail.com

***Abstract*-The elliptic curve cryptography (ECC) is a public key cryptography. The mathematical operations of ECC is defined over the elliptic curve $y^2=x^3+ax+b$, where $4a^3+27b^2 \neq 0$. Each value of the 'a' and 'b' gives a different elliptic curve. All points (x,y) which satisfies the above equation plus a point at infinity lies on the elliptic curve.**

There are certain property of elliptic curve which makes the cryptography weak. In this paper, we have proposed technique which would eliminate such weak property and will make elliptic curve cryptography more secure.

Keywords: cryptography, security, anomalous curve, discrete logarithm problem

I INTRODUCTION

Cryptography is the study of "mathematical" systems for solving two kinds of security problems: privacy and authentication [1].Two types of cryptography are present – private key cryptography and public key cryptography.

In public key cryptography, each user or the device taking part in the communication generally have a pairs of keys, a public key and a private key, and a set of operations associated with the key to do the cryptographic operations.

Only the particular user knows the private key where as the public key is distributed to all users taking part in the communication. Public key cryptography, unlike private key cryptography does not require any shared secret between communicating parties but it is much slower than private key cryptography which is main drawbacks of public key cryptography.

Elliptic curve cryptography is a variant of public key cryptography which eliminates the drawback of public cryptography. Elliptic curve $y^2=x^3+ax+b$, where $4a^3+27b^2 \neq 0$ for which each value of 'a' and 'b' gives a different elliptic curve. In ECC, public key is the point on the curve and private key is a random number. The public key is obtained by multiplying the private key with the generator point G in the curve.

One main advantage of ECC is its small size. A 160 bit key in ECC is considered to be as secured as 1024 bit key in RSA.

II BACKGROUND KNOWLEDGE

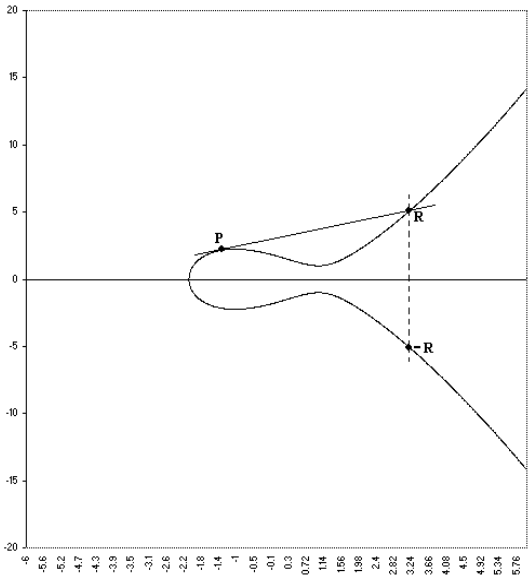
Elliptic Curves

Elliptic curves are not ellipses, instead, they are cubic curves of the form $y^2 = x^3 + ax + b$. Elliptic curves over

http://sites.google.com/site/ijcsis/
ISSN 1947-5500

R^2 (R^2 is the set R x R, where R = set of real numbers) is defined by the set of points (x, y) which satisfy the equation $y^2 = x^3 + ax + b$, along with a point $\boldsymbol{O}$, which is the point at infinity and which is the additive identity element. The curve is represented as $E(R)$.

The following figure is an elliptic curve satisfying the equation $y^2 = x^3 - 3x + 3$:-



Elliptic curve over R^2: $y^2 = x^3 - 3x + 3$

A. Elliptic Curves over Finite Fields

1) ***Elliptic Curves over F_p:*** An elliptic curve $E(F_p)$ over a finite field F_p is defined by the parameters a, b $\in$ F_p (a, b satisfy the relation $4a^3 + 27b^2 \neq 0$), consists of the set of points (x, y) $\in$ F_p, satisfying the equation $y^2 = x^3 + ax + b$. The set of points on $E(F_p)$ also include point O, which is the point at infinity and which is the identity element under addition.

2) ***Elliptic curves over F_2m:*** An elliptic curve $E(F_2m)$ over a finite field F_2m, is defined by the parameters a, b $\in$ F_2m, (a, b satisfy the relation $4a^3 + 27b^2 \neq 0$, b $\neq$ 0), consists of the set of points (x, y) $\in$ F_2m, satisfying the equation $y^2 + xy = x^3 + ax + b$. The set of points on $E(F_2m)$ also include point O, which is the point at infinity and which is the identity element under addition.

Similar to $E(F_p)$, addition is defined over $E(F_2m)$ and we can similarly verify that even $E(F_2m)$ forms an abelian group under addition.

B. Advantage of Elliptic Curve Cryptography Over RSA/DSA

The advantage of elliptic curve over the other public key systems such as RSA, DSA etc is the key strength[2]. The following table summarizes the key strength of ECC based systems in comparison to other public key schemes.

RSA/DSA Key length	ECC Key Length for Equivalent Security
1024	160
2048	224
3072	256
7680	384
15360	512

Table 1:-Comparison of the key strengths of RSA/DSA and ECC

From the table it is very clear that elliptic curves offer a comparable amount of security offered by the other popular public key for a much smaller key strength. This property of ECC has made the scheme quite popular of late.

http://sites.google.com/site/ijcsis/
ISSN 1947-5500

III ELLIPTIC CURVE DISCRETE LOGARITHM

The strength of the Elliptic Curve Cryptography lies in the Elliptic Curve Discrete Log Problem (ECDLP). The statement of ECDLP is as follows.

Let E be an elliptic curve and $P \in E$ be a point of order n. Given a point $Q \in E$ with $Q = mP$, for a certain $m \in \{2, 3, \ldots\ldots, m-2\}$.

Find the m for which the above equation holds.

When E and P are properly chosen, the ECDLP is thought to be infeasible. Note that $m = 0$, 1 and $m - 1$, Q takes the values $\boldsymbol{O}$, P and $-P$. One of the conditions is that the order of P i.e. n be large so that it is infeasible to check all the possibilities of m.

The difference between ECDLP and the Discrete Logarithm Problem (DLP) is that, DLP though a hard problem is known to have a sub exponential time solution, and the solution of the DLP can be computed faster than that to the ECDLP. This property of Elliptic curves makes it favorable for its use in cryptography.

A direct approach to determining $\#E(Fq)$ *is to compute* $z = x^3 + Ax + B$ *for each* $x \in Fq$, and then to test if z has a square root in Fq. If $z = 0$, then $(x, 0) \in E(Fq)$.

If there exists $y \in Fq$ such that $y^2 \bmod q = z$, then $(x,y),(x,-y) \in E(Fq)$, else there is no point in $E(Fq)$with x-coordinate x. So there are at most $2q + 1$ elements in the group.

A theorem of finite fields states that exactly 1⁄2 of the non-zero elements of Fq are quadratic residues. So on average, there will be approximately $q + 1$ elements in $E(Fq)$.

A. *Hasse's Theorem*

The following theorem, first proved by Helmut Hasse, told bounds on $\#E(Fq)$. Let $\#E(Fq)$ be an elliptic curve over the finite field Fq with $q = p^n$, $n \in Z+$ and p a prime. Then there exists a unique $t \in Z$ such that $\#E(Fq) = q + 1 - t$ where $|t| < 2\sqrt{q}$.[4]

B. *Reducing the problem of computing the order of curve* $\#E(Fp^n)$ *to* $\#E(Fp)$

It tells that if we can compute $\#E(Fp)$, then we can compute $\#E(F\,p^n)$ in a direct manner.Let $\#E(Fp) = p + 1 - t$.

Write $X^2 - tX + p = (X - \alpha)(X - \beta)$.

Then $\alpha^n + \beta^n \in Z$ and $\#E(F\,p^n) = p^n + 1 - (\alpha^n + \beta^n)$.

If p is a small prime, then it is easy to determine #E(Fp) by direct counting or other simple methods.

C. *Weak curves*

1) *Anomalous curve:* The curve E(Fq) is said to be anomalous if # E(Fq) = q. These curves are weak when q=p, the field characteristic.

2) *Supersingular elliptic curves:* The MOV(Menezes, Okamoto, and Vanstone) attack on elliptic curves shows that ECDLP can be reduced to the classical discrete logarithm problem on some extension field Fq^k , for some integer k (k is called the embedding degree or MOV degree). The MOV attack is only practical when k is small. For Supersingular elliptic curves $k<=6$.

3) *Prime-field anomalous curves:* If $\#E(Fp) = p$, there is polynomial algorithm solving the ECDLP by lifting the curve and points to Z.

The given properties of weak curve indicate that the order of elliptic curve plays a major role in determining whether the given curve is weak or not. The Prime-field anomalous curve and anomalous curve where the order of

http://sites.google.com/site/ijcsis/
ISSN 1947-5500

curve is a prime number can be identified with the help of Lagrange's Theorem and Hasse's Theorem.

IV PROPOSED APPROACH

A. *Lagrange's Theorem*

If G is a finite group and H is a subgroup of G, then |H| divides |G| i.e. order of subgroup H will divides the order of group G and the order of each element of the group divides the order of the group [5].

By using the above theorem an algorithm is developed to examine that the curve may have the property of Anomalous curve and Prime-field anomalous curve.

B. *Proposed Algorithm*

Step 1:- Find the solution of Equation $x^3 + ax + b=0$ which is the right hand side portion of general elliptic cuve equation $y^2 = x^3 + ax + b$.

Step 2:- Determine whether the solution of the above equation lies in the field where elliptic curve equation is defined.

Step 3:- If the solution exist in the the field then there is atleast a point (x_1, y_1) of order two i.e. $2(x_1, y_1)=\mathbf{0}$ which indicate that order of the elliptic curve can not be a prime number.

C. *Correctness of above algorithm*

If there is a point (x_1, y_1) of order two lies on the elliptic curve, then $(x_1, y_1) + (x_1, y_1) = \boldsymbol{0}$ which is point at infinity.This implifies that $(x_1, y_1) = - (x_1, y_1)$.

From the arithmetic of elliptic curve, it is known that $-(x_1, y_1)$ is a point which is mirror image of (x_1, y_1) with respect to *X*-axis. So $(x_1, y_1) = - (x_1, y_1)$ is true only when the *Y*-coordinates of (x_1, y_1) is equal to zero. It indicate that if the value of x_1 is put in the equation $x^3 + ax + b$ then it will be equal to zero.

Because of these reasons, in step 1 of the algorithm the solution of equation $x^3 + ax + b = 0$ is determined and check wether the solution lies in the field in which elliptic curve is defined.

D. *Facts derived from above algorithm*

1) The set of points $E(\boldsymbol{F}_q)$ is a finite abelian group. It is always cyclic or the product of two cyclic groups. For example the curve defined by $y^2 = x^3 - x$ over $\boldsymbol{F}_{71}$ has 72 points (71 affine points including (0,0) and one point at infinity) over this field, whose group structure is given by $Z/2Z \times Z/36Z$.

If the order of elliptic curve is prime then according to fundamental theorem of finite abelian group it is isomorphic to Z_n where *n* is prime and it is always cyclic group.

2) If the order of elliptic curve is prime then every point of elliptic curve can play the role of generator in elliptic curve cryptography.

3) The elliptic curve which has points of order *2* signifies that the order of elliptic curve is even number which reduces the range of Hasse's bound theorem which tells that order of the elliptic curve $\#E(Fq) = q + 1 - t$ *where* $|t| < 2\sqrt{q}$.

V CONCLUSION

For efficient implementation of ECC, it is important that there must be some constraints on order of the elliptic curve. In our study, we have found that there are some curves which are not suitable for elliptic curve

http://sites.google.com/site/ijcsis/
ISSN 1947-5500

cryptography because of their weak properties. These weak properties are based on the order of the elliptic curve. We have developed procedure which can identify prime-field anomalous curves which is weak and not suitable for cryptography .The proposed procedure also reduces the range of order of the elliptic curve by half.

ACKNOWLEDGMENTS

The authors would like to thank the anonymous reviewers for the valuable comments that have significantly improved the paper quality. They would also like to thanks their respective head of departments for the selfless guidance which encourage them to do this research.

REFERENCES

[1] William Stallings, *Cryptography and Network Security-Principles and Practice*, Prentice Hall Publications, Second Edition.

[2] A. K Lenstra, E.R.Verhul, *"Selecting Cryptographic key sizes"*, Nov 14 1999.

[3] Ian F. Blake, Gadiel Seroussi, and Nigel P. Smart, *Elliptic Curves in Cryptography*, London Mathematical Society Lecture Note Series, Cambridge University Press, Cambridge, 1999

[4] Advances in Elliptic Curve Cryptography (Edited by I.F. Blake, G. Seroussi and N.P. Smart). London Mathematical Society Lecture Note Series, Cambridge University Press, 2004.

[5] A Menezes, S. Vanstone, T. Okamoto, *"Reducing Elliptic Curve Logarithms to Logarithms in a Finite Field"*, IEEE transaction on Information Theory, Vol 39 (1993), 1639-1646.

[6] B.Schneier ,*Applied Cryptography* ,John Wiley and Sons, Second Edition, 1996.

[7] Alessandro Cilardo, Luigi Romano, Nicola Mazzocca and Luigi Coppolino, "*Elliptic Curve Cryptography Engineering*" PROCEEDINGS OF THE IEEE, VOL. 94, NO. 2, FEBRUARY 2006.

[8] Lawrence C. Washington , *Elliptic Curves: Number Theory and Cryptography*, 2nd edition .

http://sites.google.com/site/ijcsis/
ISSN 1947-5500

Recent Advance in Multi-Carrier Underwater Acoustic Communications

G. P. Harish
Annamalai University,
Tamilnadu, India

***Abstract*—underwater acoustic (UWA) channel is characterized as a severe multipath propagation channel due to signal reflections from the surface and the bottom of the sea and also a fast time-varying channel due to transceiver motion and medium inhomogeneities. Therefore, UWA communications have been regarded as the most challenging wireless communications. The Multi-carrier communication is a promising communication technique for future communication systems. In the past decade, much research literature focuses on deploying multi-carrier communications in UWA environments. This paper propose an overview of recent advance in multi-carrier UWA communications, which includes but not limited to Orthogonal Frequency Division Multiplexing (OFDM), Multi-input-Multi-Output (MIMO), and their related channel estimation and adaptive communications.**

Keywords- Underwater acoustic channel, OFDM, MIMO, Adaptive communications, Channel estimation.

I. INTRODUCTION

Signal propagation in underwater acoustic (UWA) environments will suffer severe multipath delay due to reflections from the sea surface and bottom. In addition, the UWA channel is a kind of fast time-varying channels due to surface wave and transceivers in motion, medium inhomogeneities and sound speed anomaly, and effect of wind-generated bubbles [1-3]. Therefore, UWA communications have been regarded as one of the most challenging wireless communication systems, especially in shallow water environments. How to achieve high data rate and reliable communications in UWA environment is one of challenging topics of wireless communications that has perplexed scientists for a long time.

Multi-carrier communications is a promising technique that could increase the system capacity and data rate significantly. Orthogonal Frequency Division Multiplexing (OFDM) is a sophisticated multi-carrier technique, which has merits of robust overcoming multipath propagation delay via cyclic prefix (CP), mitigating inter-symbol interference (ISI) and inter-channel interference (ICI). Currently, OFDM has been adopted in the 4th generation wireless communication systems, Wireless LAN network, HDTV and so on [4]. However, OFDM applications in UWA communications are very scarce [5-7].

In this paper, we propose an overview of multi-carrier communications in UWA environments. The content includes OFDM modulation-based channel estimation, OFDM and Multi-input-Multi-output (MIMO)-OFDM UWA communication systems, and their related adaptive communications.

The rest of this paper is organized as follows: Section II is the introduction of OFDM communication systems. Section III is the overview of channel estimation for UWA communications. Section IV is the overview of recent advance in multi-carrier UWA communication systems. Section V is the conclusion of this paper.

II. OFDM UWA COMMUNICATION

Figure 1 depicts eigen-ray propagation in UWA environments. Here, eigen-ray means acoustic wave path propagating from the transmitter side to the receiver side [8].

Figure 2 schematically depicts the structure of an OFDM UWA communication system. The key characteristics and principles of operation of OFDM communications include orthogonality, implementation using the FFT/IFFT algorithm, guard interval/cyclic prefix for elimination of ISI, simplified equalization, and so on [9].

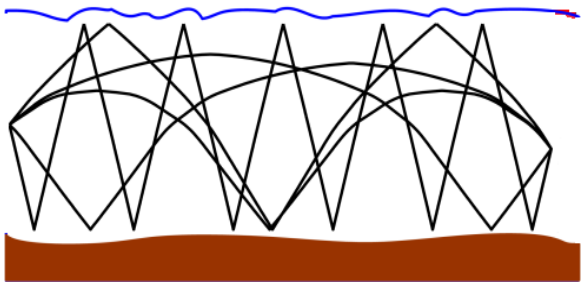


Figure 1. Schematic description of acoustic signal propagation in underwater acoustic environments

http://sites.google.com/site/ijcsis/
ISSN 1947-5500

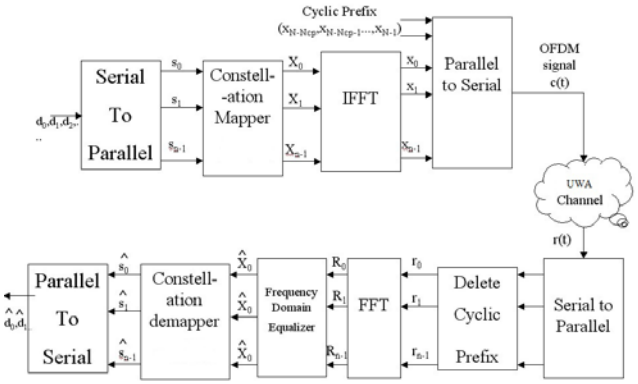


Figure 2. Schematic description of UWA OFDM communications systems

III. MULTI-CARRIER-BASED UWA CHANNEL ESTIMATION

Channel estimation techniques in UWA environments can be divided into two categories: deterministic approach and stochastic approach [10]. The deterministic approach regards the channel as a set of fixed unknown parameters to be estimated and solve a least squares estimation problem to recover the channel, while the stochastic approach exploits the second order statistics of the channels. The existing algorithms of these approaches find the proper correlation between both the time and frequency domain and linearly combine to reconstruct the channel state information (CSI) for the desired time and frequency slot. Since most of these algorithms exhibit high complexity, the applications and research of statistics approaches in UWA environments are scarce due to the difficulty of tracking fast time-varying channels. In the following of this section, we propose an overview of channel algorithms for the deterministic approach and stochatic approach, respectively.

In the deterministic approach, the channel estimator, such as Least Square (LS) and Minimum Mean Square Error (MMSE), and pilot signal are required for OFDM channel estimations. [11] proposed pilot-aided OFDM channel estimations, which involve in the block-type and comb-type pilots for OFDM systems. Authors prove that the proposed channel estimators can work effectively in both time and frequency domains for tracking fast time-varying UWA channels. [10] proposed efficient channel estimation schemes for OFDM systems in UWA environments. A robust channel estimator using pilot symbol assisted modulation (PSAM) for both single-input and single-output (SISO) and MIMO system is developed which provides excellent performance, good spectrum efficiency and manageable complexity. In [12], frequency and time correlation of the UWA channel were exploited to obtain a low-complexity adaptive channel estimation algorithm for multiple-input–multiple- output (MIMO) spatial multiplexing of independent data streams. The algorithm is coupled with non-uniform Doppler prediction and tracking, which enable decision-directed operation and reduces the overhead. In [13], the performance of three pilot patterns and their own application conditions are analyzed and compared. According to the simulation and experiment results, it is concluded that scatter pilot pattern is very suitable for OFDM system for underwater acoustic communications. Besides, the other deterministic algorithms with significant performance, they can be found in [14-16].

In the stochastic approach, [17] considered UWA channel estimation based on sparse recovery using the recently developed homotopy algorithm. The UWA communication system under consideration employs OFDM and receiver preprocessing to compensate for the Doppler Effect before channel estimation. [18] provided a novel UWA Channel estimation and Simulator based on measured scattering functions.

In addition to these two categories, there is much literature engages in establishing effectively channel estimation methods for OFDM UWA communications. [19] investigated two methods for estimating the matched signal transformations caused by time-varying UWA channels in OFDM communication systems. The first channel estimation method is based on discretizing the wideband spreading function time-scale representation of the channel output using the Mellin transform. The second method is based on extracting the time-scale features of distinct ray paths in the received signal using a modified matching pursuit decomposition algorithm.

IV. RECENT ADVANCE IN MULTI-CARRIER UWACOMMUNICATIONS

A. OFDM UWA Communications

We discuss several important issues of OFDM UWA communications. Due to unique properties of UWA channels, OFDM UWA communication systems have many different points compared with radio frequency OFDM communications. [20] applied OFDM to realize parallel transmission of spread spectrum signal in UWA communications, so as to provide robust acoustic links or long distance communication abilities. The traditional CP-based OFDM communications using a overlap-add method have a bad performances when channel is severe frequency-selective, especially with channel nulls, which is often encountered in UWA channels, [21] utilized zero-padding (ZP)-OFDM channel equalization on the premise of the channel transfer matrix is Toeplitz matrix, Monte-Carlo simulation proved that this method has a better performance than CP-OFDM, and has a good application prospect for UWA communications. [22] presented a desirable property of OFDM that one signal design can be easily scaled to fit into different transmission bandwidths with negligible changes on the receiver.

Doppler Shift is an important factor that affects the performance of UWA communication systems. Therefore, how to overcome the Doppler Shift problem in OFDM UWA communications becomes a challenging issue. [23] focused on ZP-OFDM to minimize the transmission power. In addition, authors treated the channel as having a common Doppler scaling factor on all propagation paths, and propose a two-step approach to mitigating the Doppler effect: (1) non-uniform

http://sites.google.com/site/ijcsis/
ISSN 1947-5500

Doppler compensation via resampling that converts a "wideband" problem into a "narrowband" problem and (2) high-resolution uniform compensation of the residual Doppler. [24] studied the performance of OFDM over UWA multipath channels with different Doppler scales on different paths. [25] treated the channel as having a common Doppler scaling factor on all propagation paths, and propose a novel approach to mitigating the Doppler effects in OFDM UWA communication systems.

Mitigation of ICI and ISI of OFDM UWA communication systems is another challenging issue for achieving high data rate and reliable communications. [26] focused on CP-OFDM over time-varying UWA channels. To cope with the ICI that arises at the receiver side because of the time variations in the channel, authors considered two ICI-mitigation techniques. In the first scheme, the ICI coefficients are explicitly estimated, and minimum mean square error linear equalization based on such estimates is performed. In the second approach, no explicit ICI estimation is performed, and detection is based on an adaptive decision-feedback equalizer applied in the frequency domain across adjacent subcarriers.

Real implementations and performance analysis of OFDM UWA communication systems have been investigated by many researchers. [27] designed and implemented the OFDM signal transmitter with FPGA (field programmable gate array) and DSPs (digital signal processor, ADSP-TS101). [28-28] analyzed the performance of capacity criterion-based OFDM UWA communications. Above all, [29] derived bounds to the channel capacity of OFDM systems over the UWA fading channel as a function of the distance between the transmitter and the receiver. The upper bound is obtained under perfect CSI at the receiver. The lower bound is obtained assuming the input is drawn from phase-shift keying (PSK) constellation which results in non-Gaussian distribution of the output signal and no CSI.

B. MIMO-OFDM UWA Communications

The MIMO-OFDM scheme is one kind of more advance communication technique for UWA communications. MIMO-OFDM could further increase the system capacity and data rate over the bandwidth limited channels. [30-31] presented a MIMO system design, where spatial multiplexing is applied with OFDM signals. The proposed receiver works on a block-by-block basis, where null subcarriers are used for Doppler compensation, pilot subcarriers are used for channel estimation, and a MIMO detector consisting of a hybrid use of successive interference cancellation. [32-33] provided further results of MIMO-OFDM UWA Communications. [34] analyzed MIMO-OFDM communications for shallow water environments, which is more challenging than normal UWA communication systems.

C. Adaptive Multi-Carrier UWA Communications

UWA communications possess properties of several channel fading and limited bandwidth resource. Therefore, adaptive techniques are more valuable to be adopted in UWA communications, especially for shallow water environments. In order to achieve adaptive signal transmission, information of the receiver side is indispensable. However, due to limited bandwidth resource, traditional perfect feedback techniques used in radio frequency wireless systems become impractical for UWA communications. [35-38] involved limited feedback techniques in UWA OFDM communications for the first time, which makes adaptive signal propagation and resource allocation for complicated UWA environments possible. This innovation can be regarded as an important breakthrough for UWA communications, which could significantly increase the system performance while save communication resource simultaneously. Furthermore, [39] analyzed the minimum BER-based performance of adaptive OFDM UWA communications with limited feedback. Other adaptive multi-carrier UWA communication techniques can be found in [40-41].

V. Conclusion

This paper provided an overview of multi-carrier communications in UWA environments. Future research can focus on multi-carrier techniques together with other advanced wireless communication techniques for UWA communications, such as OFDM with cooperative transmission, and OFDM with cognitive radio. Definitely, these techniques will significantly improve the performance of UWA communication systems.

References

[1] W. Yang and T. C. Yang, “High-Frequency Channel Characterization for M-ary Frequency-Shift-Keying Underwater Acoustic Communications”, Journal of Acoustical Society of America, vol. 120, no. 5, pp. 2615-2626, August 2006

[2] T. C. Yang, "Temporal Coherence of Acoustic Rays and Modes Using the Path Integral Approach”, Joural of Acoustical Society of America, vol.131, no. 6, pp. 1716-1722, June 2012

[3] T. C. Yang, “Properties of Underwater Acoustic Communication Channels in Shallow Water”, Journal of Acoustical Society of America, vol.131, no. 129, pp. 129-145

[4] http://en.wikipedia.org/wiki/Orthogonal_frequency-division_multiplexing

[5] D. Wang, R. Xu, S. Zheng, F. Xu, X. Hu and H. Liu, “Research on Based-Band OFDM Underwater Acoustic Communication System”, ICISE conference, 2703-2706, 2009

[6] L. Zhang, X. Xu, H. Sun and Y. Chen, “Performance Analysis of IRA Codes for Underwater Acoustic OFDM Communication System”, WiCom Conference, pp.1-4, 2009

[7] P. Kumar, “DCT Based OFDM for Underwater Acoustic Communication”, RAIT Conference, pp.170-176, 2012

[8] S. Byun, S. Kim, Y. Lim, and W. Seong,”Time-Varying Underwater Acoustic Channel Modeling for Moving Platform”, IEEE Oceans Conference, pp. 1-4, 2007

[9] A. A. Hutter, “Design of OFDM Systems for Frequency-Selective and Time-Variant Channels”, International Seminar on Broadband Communications, pp.1-6, 2002

[10] D. N. Liu, S. Yerramail, and U. Mitra, “On Efficient Channel Estimation for Underwater Acoustic OFDM Systems”, ACM WUWNet conference, pp. 1-8, 2009

[11] X. Huang and V. B. Lawrence, “OFDM with Pilot Aided Channel Estimation for Time-Varying Shallow Water Acoustic Channels”, IEEE CMC conference, pp.442-446, 2010

http://sites.google.com/site/ijcsis/
ISSN 1947-5500

[12] B. Li, S. Zhou, M. Stojanovic, and L. Freitag, "Multicarrier Communication over Underwater Acoustic Channels with Nonuniform Doppler Shift", IEEE Journal of Oceanic Engineering, vol.33,no.2, pp.198-209, 2008.

[13] Y. Lv and C. Zheng, "A study of Channel Estimation in OFDM System Based on a Single Vector Sensor for Underwater Acoustic Communications", WiCom conference, pp.1-4, 2008

[14] M. Stojanovic," OFDM for Underwater Acoustic Communications: Adaptive Synchronizations and Sparse Channel Estimation", IEEE ICASSP conference, pp. 5288-5291, 2008

[15] K. Grythe and J. E. Hakegard, "Non-Perfect Channel Estimation in OFDM-MIMO Based Underwater Communication", pp.1-9, 2009.

[16] K. Sunwoo, "Angle-Domain Frequency-Selective Sparse Channel Estimation for Underwater MIMO-OFDM System", IEEE Communication Letters, vol. 16, no. 5, pp. 685-687, 2012

[17] C. Qi, X. Wang and L. Wu, "Underwater acoustic channel estimation based on sparse recovery algorithms", IET Signal Processing, vol. 5., no. 8, pp.739-747, Dec. 2011

[18] J. Zhang, J. Cross, Y. R. Zheng, "Statistical Channel Modeling of Wireless Shallow Water Acoustic Communications from Experiment Data", IEEE Milcom conference, pp.105, 2011

[19] N. F. Josso, J. J. Zhang, D. Feronani, A. Papandreou-Suppappola, and T. M., Duman,"Time-Varying Wideband Underwater Acoustic Channel Estimation for OFDM Communications", IEEE ICASSP conference, pp.1-4, 2010

[20] J. Huang, C. He, Q. Zhang and H. Jing, "A Novel Spread Spectrum OFDM Underwater Acoustic Communication", IET International Conference on Wireless Mobile and Multimedia Networks, pp.1-4, 2006

[21] E. Song, X. Xu, G. Qiao, J. Su, " Study on ZP-OFDM for Underwater Acoustic Communication", IEEE International Conference on Neural Networks and Signal Processing, pp. 299-302, 2008.

[22] B. Li, S. Zhou and J. Huang, "Scalable OFDM design for underwater acoustic communications", IEEE ICASSP conference, pp.5304-5307, 2008.

[23] B. Li, S. Zhou, M. Stojanovic, and L. Freitag, "Multicarrier Communication over Underwater Acoustic Channels with Nonuniform Doppler Shift", IEEE Journal of Oceanic Engineering, vol.33,no.2, pp.198-209, 2008.

[24] S. Mason, C. Berger, S. Zhou, K. Ball, L. Freitag and P. Willett, "An OFDM Design for Underwater Acoustic Channels with Doppler Spread", IEEE 5th DSP/SPE conference, pp.138-143, 2009.

[25] T. Guo, D. Zhao, and Z. Zhang, "Doppler Estimation and Compensation for Underwater Acoustic OFDM Systems", IEEE CSQRWC, pp.863-867, 2011.

[26] K. Tu, D. Fertonani, T. M. Duman, M. Stojanovic, J. G. Proakis, and P. Hursky, "Mitigation of Intercarrier-Interference for OFDM Over Time-Varying Underwater Acoustic Channels", IEEE Journal of Oceanic Engineering, vol. 36, no.2, pp.156-171, 2011

[27] K. Lei, Z. Yan, J. Han and J. Huang, "Design and Implementation of Underwater OFDM Acoustic Communication Transmitter", IEEE ICLIP conference, pp. 609-613, 2008

[28] B. Srinivasan and V. Rodoplu, "Capacity of Underwater Acoustic OFDM Cellular Networks", IEEE Oceans Conference, pp.1-10, 2010

[29] C. Polprasert, J. A. Ritcey, and M. Stojanovic, " Capacity of OFDM Systems Over Fading Underwater Acoustic Channels", IEEE Journal of Oceanic Engineering, vol. 36, no. 4, pp.514-524, 2011

[30] B. Li, J. Huang, S. Zhou, K. Ball, M. Stojanovic, L. Freitag and P. Willett, "MIMO-OFDM for High-Rate Underwater Acoustic Communications", IEEE Journal of Oceanic Engineering, vol.34, no. 4, pp. 634-644, 2009.

[31] B. Li, J. Huang, S. Zhou, K. Ball, M. Stojanovic, L. Freitag and P. Willett, "Further Results on High-Rate MIMO-OFDM Underwater Acoustic Communications", IEEE Oceans Conference, pp. 1-6, 2008.

[32] J. Huang, S. Zhou, J. Huang, J. Preisig, L. Freitag, and P. Willett, "Progressive MIMO-OFDM Reception over Time-Varying Underwater Acoustic Channels", ASILOMAR conference, pp. 1324-1329, 2010

[33] M. Stojanovic, "MIMO-OFDM over Underwater Acoustic Channels", ASILOMAR conference, pp.605-609, 2009.

[34] P. Bouvet and A. Loussert, "An Analysis of MIMO-OFDM for Shallow Water Acoustic Communications", IEEE Oceans Conference, pp.1-5, 2011

[35] X. Huang, "Capacity criterion-based power loading for underwater acoustic OFDM system with limited feedback", IEEE WCNIS conference, pp.54-58, 2010.

[36] X. Huang and V. B. Lawrence, "Capacity Criterion-Based Bit and Power Loading for Shallow Water Acoustic OFDM System with Limited Feedback", IEEE 73rd Vehicular Technology Conference, pp.1-5, 2011

[37] X. Huang and V. B. Lawrence, "Bandwidth-Efficient Bit and Power Loading for Underwater Acoustic OFDM Communication System with Limited Feedback", IEEE 73rd Vehicular Technology Conference, pp.1-5, 2011

[38] X. Huang and V. B. Lawrence, "Effect of wind-generated bubbles on OFDM power loading for time-varying shallow water acoustic channels with limited feedback", IEEE Oceans Conference, pp.1-6, 2011

[39] A. Radosevic, T. M. Duman, J. G. Proakis and M. Stojanovic, "Adaptive OFDM for Underwater Acoustic Channels with Limited Feedback", ASILOMAR conference, pp.975-980, 2011

[40] R. F. Ormondroyd, "A Robust Underwater Acoustic Communication System Using OFDM-MIMO", IEEE Oceans Conference, pp.1-6, 2007

[41] Y. Lei, L. Zhou and M. Yu, "Adaptive Bit Loading Algorithm for OFDM Underwater Acoustic Communication System", ICEOE conference, pp.350-352, 2011

http://sites.google.com/site/ijcsis/
ISSN 1947-5500

Decreasing defect rate of test cases by designing and analysis for recursive modules of a program structure: Improvement in test cases

Muhammad Javed[1], Bashir Ahmad[1] , Zaffar Abbas[1], Allah Nawaz[1] , Muhammad Ali Abid[1] , Ihsan Ullah[1]
[1]*Institute of Computing and Information Technology Gomal University, D.I.Khan, Pakistan*

Abstract---***Designing and analysis of test cases is a challenging tasks for tester roles especially those who are related to test the structure of program. Recently, Programmers are showing valuable trend towards the implementation of recursive modules in a program structure. In testing phase of software development life cycle, test cases help the tester to test the structure and flow of program. The implementation of well designed test cases for a program leads to reduce the defect rate and efforts needed for corrective maintenance. In this paper, author proposed a strategy to design and analyze the test cases for a program structure of recursive modules. This strategy will definitely leads to validation of program structure besides reducing the defect rate and corrective maintenance efforts.***

Index Term—Test cases, Recursive module, Black-box, White-box, corrective maintenance, defect rate.

I. Introduction

Testing phase of software development life cycle lead to the quality of software products and it depend on the strategies which are followed by tester role. The most commonly used methods of testing are black-box and white-box testing [1]. In black-box tester examine the fundamentals aspects of software; while in white-box tester examine the internal procedure detail of the system components such path testing and loop testing. During white-box testing test cases can be generated either manual or through automated tool to check the working of software. A test case is a set of conditions or variables which are included in the working of software[5,7]. The focus of this paper is to design and analyze the generation of test cases for recursive modules in programming language. Here author's proposed a strategy which helps to reduce the defect rate and corrective maintenance efforts.

II. Recursive Modules

In programming language structure recursive modules are those routines which called itself during execution of program and they can consider as central idea of computer science [6, 8]. There are two factors which are relevant to recursive modules. First is the base case used to end the calling of recursive module and second is to break the current domain of data into sub domains and this will remain continue till base case satisfied [10,11]. Recursive modules are classified into linear, mutual, binary and N-Ary Types.

III. White box testing

White box testing is the process to test the implementation of a system. It consist of analysis of data flow, control flow, information flow , coding practice and exception handling within the system to ensure correct software behavior. White box testing can be perform by tester any time after coding but it will be good practice to do it with unit testing. White box testing is used with unit, integration and regression testing. In white box testing method tester role can perform the following activities[3].

- It defines the test strategy and activities.
- It develop new test plan on the base of selected strategy.
- It creates an environment for test case execution.
- It executes the test cases and prepared reports.

The main types of white box testing are static and dynamic analysis, branch coverage, security testing and mutation testing. Selection of skilled tester and bit of code to remove error are considered as important challenge in white box testing [9].

In software project the success of testing depend on the test cases used. To reduce the turn around time, defect rate and project duration it is important to design an effective set of test cases that enables detection of maximum number of errors [12].

IV. Flow Graph Notation (FGN)

In white-box testing Flow Graph Notation (FGN) is a used to represent the program control structure. It is just like flowchart and comprises on circle and edges. Each circle, called a flow graph node, represents one or more procedural statements and edges represent the flow of control. An edge must terminate at a node, even if the node does not represent any procedural statements. Areas bounded by edges and nodes are called regions. When counting regions, we include the area outside the graph as a region.

V. Analysis of Test Cases for Recursive Modules

To represents the analysis and design process of recursive modules an example in C++ language is taken as shown in Fig-1. In this example two recursive modules/functions are used named as "Factorial" and "SumofFact". Following steps are used to represent the working of C++ program shown in Fig-1.

http://sites.google.com/site/ijcsis/
ISSN 1947-5500

```
#include<iostream.h>
#include<stdio.h>
#include<conio.h>
int Factorial(int);
int SumofFact(int);
int sum=0,temp;
void main()
{
  clrscr();
  int number;
  cout<<"Enter Number to find factorial";
  cin>>number;
  cout<<"Factorial of number is = " << Factorial(number);
  cout<<"\nSum of factorials of all number from 1 to n"<< SumofFact(number);
  getch();
}
int Factorial(int n)
{
 if(n<1)
   return(1);
 else
   return(n*Factorial(n-1));
}
int SumofFact(int n)
{

if(n>0)
 { sum=sum+Factorial(n);
 temp=SumofFact(n-1);}
else
  return(sum);
}
```

Fig-1. C++ Program including recursive modules/functions

- Firstly a number is read in the main module/function of program.
- Secondly a recursive module named “Factorial” is called from main function to find the factorial of entered number. If number is 4 then result of “Factorial” function will be 24 i.e. 4! = 24.
- In third step another recursive module named “SumofFact” is called to add the sum of factorial of all numbers ranges from 1 to entered number. If number is 4 then result of this function will be 1!+2!+3!+4!=33

To analyze the complexity of program (shown in Fig-1) a Flow Graph Notation is drawn which is shown in fig-2. This FGN represent the all paths which can be used to analyze and design the test case for program. As there two factors, which are related with recursive module, first is the base condition which is applied to end the calling of recursive modules and second factor is relevant to division of domain of data for recursive module into sub domains. The complexity of recursive module calling can be analyzed with respect to two aspects.

1. Calling of a recursive module from any other module which is not recursive in nature.
2. Calling of a recursive module from any other module which is recursive in nature.

http://sites.google.com/site/ijcsis/
ISSN 1947-5500

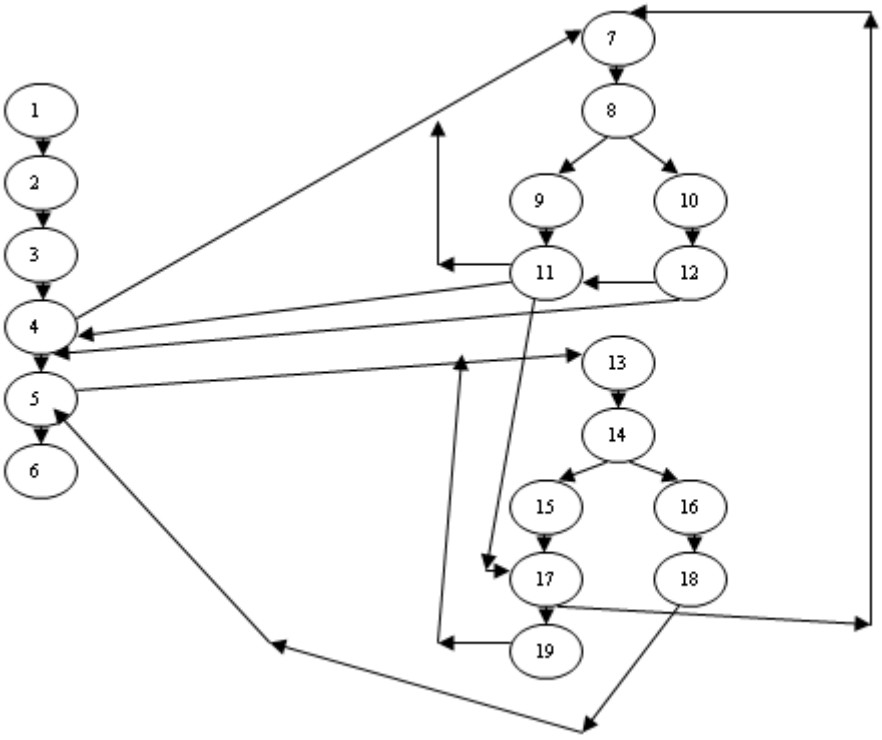


Fig-2. Flow Graph Notation for C++ Program of recursive modules

The complexity of program will be high for second aspect as compared to first. The program shown in Fig-1 represents the both aspect of calling the recursive module. The first aspect is represented through calling of "Factorial" recursive module/function and second is through calling of "SumofFact". In "SumofFact" recursive module/function "Fact" is again called. This process leads to increase the complexity of program.

Analysis and Designing of Test cases and test data for the first Aspect:

The first aspect shows the calling of recursive module/function from another function which is not recursive in nature. If we omit the "SumofFact" recursive function from program shown in Fig-1 and its calling from main module. Then there be will only two possible path to represent the execution of "Factorial" recursive module.

Path-1.

1-2-3-4-7-8-10-12-5

Path-2.

1-2-3-4-**7-8-9-11**-7-8-10-12-**11**-5

The first path represents the execution of statements of calling and called module in sequence. Which show the recursive module "Factorial" is called only one time from "main" module and it is not called by itself. The test case for this path will be n<1 and test data for this test case may be 0 or any negative number.

The second path represent that recursive module is called many time depend on the domain of data, this is shown in highlighted part of path i.e. 7-8-9-11. When the recursive module is called by itself last time then base condition will be executed which is shown in 7-8-10-12 part of second path. After that control will be transfer by recursive module to itself, this is shown in another highlighted part of second path i.e. 11. At the end control will be transfer back to the "main" calling module of recursive module. The test case for this path will be n>=1 and test data may be any positive value. If one recursive module is called many times from "main" calling module then same two paths will be used except the nodes of FGN will be increases. It is clear from this analysis that test case and test data will remain same whether you will called a recursive module one or more than one time.

Analysis and Designing of Test cases and test data for second Aspect:

The second aspect shows the calling of recursive module/function from another function which is recursive in nature. According to program of Fig-1 "SumofFact" is the calling module of "Factorial" recursive module and "SumofFact" itself is recursive module. To analyze the test cases for this aspect firstly omit the node 4 from FGN of "main" module. This will show that "Factorial" recursive module will not called from "main" module. There be will only two possible path to represent the execution of "SumofFact" and "Factorial" recursive modules.

Path-1.

1-2-3-5-13-14-16-18-6

Path-2.

1-2-3-5-**13-14-15-17-7-8-9-11-19**-13-14-16-18-6

The first path represents the execution of statements of calling (i.e. "main" module) and called module(i.e "SumofFact") in

http://sites.google.com/site/ijcsis/
ISSN 1947-5500

sequence. In this path execution of "Factorial" recursive module is not shown because here the base condition of "SumofFact" is executed and it not called itself. The test case for this path will be n<0 and test data may be 0 or any negative number. The second path represents the more than one time execution of "SumofFact" and "Factorial" modules. The part of second path i.e. 13-14-15-17-7-8-9-11-19 as a whole represents the recursive execution of both modules. In this part of second path i.e. 13-14-15-17 represents the execution of "SumofFact" and calling of "Factorial" recursive modules. Moreover, in this part of second path i.e. 7-8-9-11 represents the execution of "Factorial" recursive module and returning control back to the node 19 of FGN. This node also represents the calling of "SumofFact" recursive module i.e. same process will be repeated till the test case n<0 is satisfied. The test case for this path will be n>0 and test data may be any positive number. Moreover, from this analysis it is clear that first path eliminate the execution of base condition of "Factorial" recursive module, but it will not true for all cases. This is also illustrating here that during analysis and designing of test cases, some test cases can not show the execution of some part of a recursive module. So there is need to be more care during analysis and designing of test cases of recursive modules especially when a recursive module call another recursive module. If tester role will not care about it then it can leads to increase the defect rate and corrective maintenance efforts. Besides caring of tester role in analysis of recursive modules, it must care about the levels. If level to call one recursive module within another recursive module is increases then complexity of program will high and it will leads towards increases in defect rates.

VI. CONCLUSION

During white-box testing process the use of FGN and deriving path are the basis steps to analyze and design the test cases and test data. In this paper authors adopt a strategy to analyze and design the test cases for recursive modules, which are considered as important paradigm in programming language. After analysis and designing process of test case authors known that some part of the recursive modules can not be implemented through test case which can increase the defect rate and corrective maintenance efforts.

REFERENCES

[1] L.S. Chin, D.J. Worth, and C. Greenough, "A Survey of Software Testing Tools for Computational Science", RAL-TR-2007-010, June 29, 2007.
[2] M.Prasanna at Al, "a survey on automatic test case generation", Academic Open Internet Journal, Volume 15, 2005.
[3] Vinod Dandoti, "White Box Testing: An Overview", 2005.
[4] Prof Marsha Chechik, "Test Generation using Model Checking" 2000.
[5] ADVENT , Advance InfoSystems LLC, "Pre-Packaged Test Cases", 2008.
[6] Andrew Myers, "Recursive names and modules", 18 February 2009.
[7] Baikuntha, Pragyan Nanda and Durga Prasad, "A Novel Approach for Scenario-Based Test Case Generation", 2008 IEEE.
[8] Keiko Nakata and Jacques Garrigue, "Recursive Modules for Programming", 2006/9/26.
[9] Laurie Williams," White-Box Testing", 2006.
[10] http://www.allisons.org/11/AlgDS/Recn/
[11]http://en.wikipedia.org/wiki/Recursion_(Computer_Science)
[12]http://www.edistalearning.com/Demo_Courses/SE500/mod6/les02/l02_000_000.htm.

Text Hiding Based on True Color Image Classification

Shahd Abdul-Rhman Hasso
Department of Computer Science, College of Computer Sciences and Math.,
University of Mosul / Mosul, Iraq

***Abstract*— In this work a new approach was built to apply k-means algorithm on true colored images (24bit images) which are usually treated by researchers as three image (RGB) that are classified to 15 class maximum only. We find the true image as 24 bit and classify it to more than 15 classes. As we know k-means algorithm classify images to many independent classes or features and we could increase the class number therefore we could hide information in the classes or features that have minimum number of pixels which are considered unimportant features and reconstruct the images.**

Correlation factor and Signal to Noise Ratio were used to measure the work and the results seems that by increasing the image resolution the effect of removing minimum features is decreased.

The MATLAB 2010 application language was used to build the algorithms which are able to allocate huge matrices especially im image processing.

Keywords-component; k-means clustering, steganography, data hiding; True color images.

INTRODUCTION

Secret communication achieved by hiding the existence of a message is known as steganography, derived from the Greek words "stegano", meaning covered and "graphy", meaning to write. In the fifteenth century, the Italian scientist Giovanni Porta described how to conceal a message within a hard-boiled EGG by making an ink from a mixture of one ounce of aluminum and a pint of vinegar, and then using it to write on the shell.

The solution penetrates the porous shell, and leaves a message on the surface of the hardened EGG albumen, which can be read only when the shell is removed new technologies were developed which could pass more information and be even less conspicuous. The Germans developed microdot technology which FBI Director J. Edgar Hoover referred to as "the enemy's masterpiece of espionage." Microdots are photographs the size of a printed period having the clarity of standard-sized typewritten pages. The first microdots were discovered masquerading as a period on a typed envelope carried by a German agent in 1941. The message was not hidden, nor encrypted. It was just so small as to not draw attention to itself (for a while). Besides being so small, microdots permitted the transmission of large amounts of data including drawings and photographs [1].

For every step steganography has taken to hide the data over the past 1500 years, mankind has worked hard to find the hidden messages. With today's computer steganographics, finding and decoding the hidden messages have become more complicated. Currently, steganalysts are working hard to detect the hidden messages within images, documents, and sound. Steganalysis starts with suspected data files. The steganalyst uses forensic statistician information to help reduce the number of files. The analyst then compares the questionable data files to similar data files. The similarity is based on the same digital camera or digital audio device. 16 The analyst is looking at visual detection (jpeg, bmp, gif, etc.), audible detection (wav, mpeg, etc.), statistical detection (changes in patterns of pixels or Least Significant Bit) or histogram analysis, and structural detection (view file properties/content, size difference, date/time difference, contents – modifications, checksum).17 Once steganography is detected, and the information is extracted, it may still be encoded. At this point, cryptanalysis techniques may be applied. Steganalysts have just started their battle against the hidden data. Much more must be done to detect the dangerous data hidden behind the innocent looking pictures [1].

It is important to understand that steganography is very different than cryptography and the two are often confused. With cryptography, encryption is the process of obscuring information to make it unreadable without some type of special knowledge. In this case the message is not concealed just scrambled or obscured [2].

The obvious advantage of steganography over cryptography is that messages do not attract any attention. A coded message that is unhidden, no matter how strong the encryption, will arouse suspicion and may in itself be problematic. For example, in some countries encryption is illegal. Stego may even be mixed with encryption so the carrier file actually carries a message that is encrypted. So even if intercepted, another barrier is presented in trying to break the encryption [2].

In general there are four steganography basic methods as follows:

1) *text hiding*
2) *voice hiding*
3) *video hiding*
4) *Image hiding*

In this work, the image hiding is applied.

IMAGE PROCESSING

Image processing aim is to build applications that are able to understand the content of images as understood by human. Where it is possible to take several forms of image data such

http://sites.google.com/site/ijcsis/
ISSN 1947-5500

as images of Video, scenes from several cameras, several dimensions of data taken from a medical imaging device. Some examples of applications of image processing [3]:

- Application is able to identify the objects or persons within the image
- Applications of automatic control (the robot and motor vehicles).
- Build models of objects or the environment (industrial inspection, medical image analysis).
- Application is able to follow a moving object within an image
- Application is able to see the third dimension from one or more two-dimensional image (or from an image and a moving laser light) [3].

COLOR CONCEPTS IN DIGITAL IMAGES

Form the color model red, green and blue (RGB), a color model combines the lights red, green and blue with each other in different ways to generate a wide range of colors.
The main objective of the RGB color model is sense, generate and display the images in electronic devices, such as computer screens [4]. The digital image is divided based on the colors into three main types:

A. Binary Images:

The binary image is the simplest basic types of digital images; each element of the image represents the one of value two values that is displayed as white and black. Numerically, the two values are represent by "1" for white and "0" for black and stored in a two-dimensional matrix of zeros and ones. The binary image is also called several names as Monochrome Image, 1 Bit Image Pixel or Black and White Image because it takes a binary representation for each point [4].

B. The Gray Level Images

Gray Level images Contain lighting information only, with no color information. This type is commonly used in digital image processing. The colors in this type of images are shades of grayscale, as the gray color is produced when the values of intensity of the colors red, green and blue are equal in the space of RGB. The number of bits used for each pixel of light determines the number of lighting levels, and ideal image data contains (8Bit / Pixel), it is allow us to have 0-255 of the different gray gradients [4]. The grayscale images are commonly used due to the fact that a lot of display devices and the acquisition systems can process images of (8 Bit) Moreover, the grayscale images are easy for many tasks, and there is no need to use of harder and more complex processes as is the case of color images [3].

II. The Digital Color images

The digital Color images represents by a separate values of the intensity of the three main colors (RED, GREEN, BLUE), because the color of each pixel is set at a gathering of those colors intensities. For storing 24 Bits color images, each color is represented by 8 Bits. This produces 16 million potential. The pixel of color (0, 0, 0) was black and the pixel of the contents of color (255, 255, 255) was white, so this type of image is known as the (24-Bit Color Image). It is efficiently cover the full range of colors that understood by the human eye but there are some disadvantages in using this type of images, where it needs more memory and takes longer to storage [4].
The 24-bit color images are also called true color images because each color values is presented fairly the on-screen by the real number of bit (8 bits) for each color of the three primary colors (red blue and green). These images represent the matrix as follows:

R	G	B	R	G	B	R	G	B	.....
R	G	B	R	G	B	R	G	B	....(1)
R	G	B	R	G	B	R	G	B	.....

In other words, each pixel is a 24-bit number (0 - 16,777,215) and the most important characteristics of these images to be high precision and homogeneity of the colors is very large, making it a clear vision [5].
But at the same time, these images contain unimportant information or features that could be canceled and deleted without affecting the image.

III. K-MEANS CLASSIFICATION TECHNIQUE

When we think of hiding in a text within images, you will surely need certain pixels to store text; these pixels must have certain characteristics collected within a certain type. Since we want to remove these pixels of commonality surely the characteristics must be unimportant so that when it is changed, it is not affected or at least the effect will not be visible.
Based on this, we need a certain algorithm to divide the image to a number of varieties. The classification algorithms could be used to do that. The K-Means clustering algorithm is a high-quality classification algorithm, with a definite result in access to the target that is required. The K-Means clustering algorithm has been developed in 1967 by J. MacQueen and then in 1975 was developed by both J. A. Hartigan and M. A. Wong. This algorithm is based on the classification of objects depending on the specific properties of this object.

The mathematical representation of k-means algorithm is as follows [5] [6]:

Step 1: determine the number of classes (the value of k).

Step 2: Choose the centers Zi of these classes. In this work, a new class selection was proposed that is by calculation the minimum and maximum of an image and selection the median values in between ending with the number of applied classes. Figure (1) shows the class selection technique

http://sites.google.com/site/ijcsis/
ISSN 1947-5500

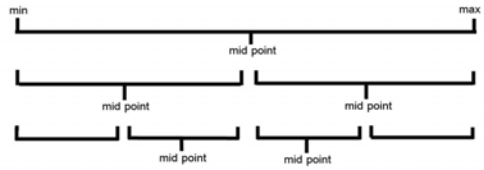


Figure (1): The class selection technique

Step 3: Calculate the Euclidean distance (Ed) between image pixels and centers of classes according to the following equation:

$$Ed = \left\| Z_j(n) - X \right\|$$

Where k represents the number of classes and j = 1, 2, 3... k
X is the image pixel to be classified.

Z is the center of classes, n represents the iteration number

Step 4: Set the image pixel to a group class Sj(n) of $\left\| Z\ (n) - X \right\|$ minimum distance.

Step 5: Calculate new centers for each class and it calculates average of pixel within each class, according to the following equation:

$$Zj(n) = 1/N_j \sum Xi$$

Where N_j represents the number of pixel in the set S_j
Step 6: Compare the old centers Zi (n) with the new centers Zi (n +1).

For the current iteration if different centers and at least one re-calculation algorithm, starting from the third step, otherwise this algorithm stops, figure (2) shows the block diagram of k-means clustering algorithm.

The K_Means algorithm is widely used in many applications not only to classify and organize data, but also it is useful in pattern recognition and information retrieval, identification of sound, the words of the speaker and Data Mining [5].

One of the disadvantages of this algorithm is that it takes a long execution time and in the phase redundancy to correct centers varieties either in terms of accuracy it is the best among the algorithms, depending on the mechanism of which is the identification of centers of classes since the update center class is not until after the testing of all types existing. [6].

IV. THE PROPOSED METHOD

A. *Hiding Methid*

Step 1: Read the color image with 24-bit three-dimensional matrix. The first dimension is the indicator the three primary colors and the second dimension and third the image size in pixel (raw X column). Also read the text file.

Step 2: Convert the image three-dimensional to two-dimensions for obtaining the (24 bit) value as it is, according to the following equation:

*Val_24bit = Val_Red + Val_Green * 256 + Val_Blue * 65536 for each image pixel*

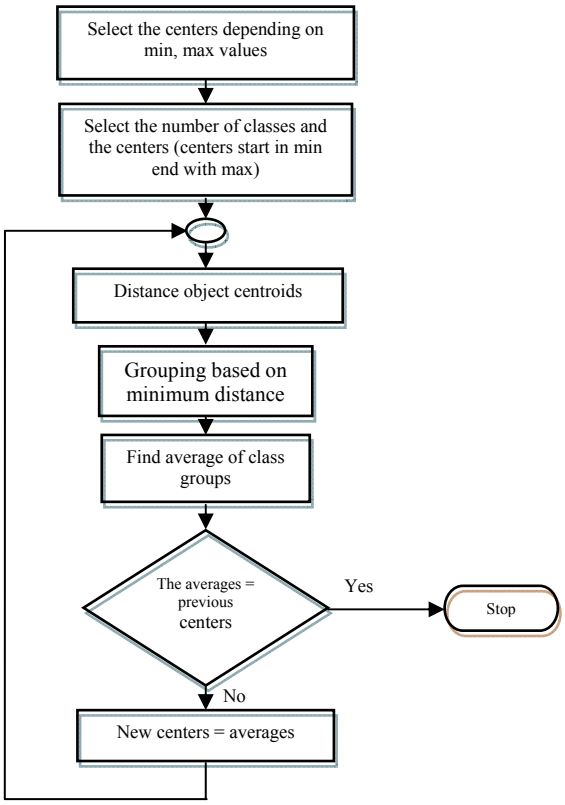


Figure (2) The block diagram of k-means algorithm

Step 3: Apply the K-means algorithm on the image storing the coordinates of each pixel classified

Step 4: Apply sorting depending on the number of pixel on classes. The minimum number of pixel (i.e., smallest class) has a few important features (ineffective features).

Step 5: Hide the data (text) in the smallest classes in its pixel coordinates.

Step 6: Convert 24-bit values to the values of the three basic colors, according to the following equation:

Val_Red = Val_24bit & 256;

Val_Green = Val_24bit& 65280) / 256;

Val_Blue = Val_24bit& 16711680) / 65536;

For each image pixel;

Figure (3), shows the flow chart of the hiding stage in the proposed method

B. *UnHiding Method*

Step 1: Repeat the same first four steps in hiding.

Step 2: Read the stego image and convert to 24 bit.

Step 3: read the data (text) in the smallest classes in its pixel coordinates.

http://sites.google.com/site/ijcsis/
ISSN 1947-5500

Step 4: Convert 24-bit values to the values of the three basic colors.

Figure (4), shows the flow chart of the unhiding stage in the proposed method

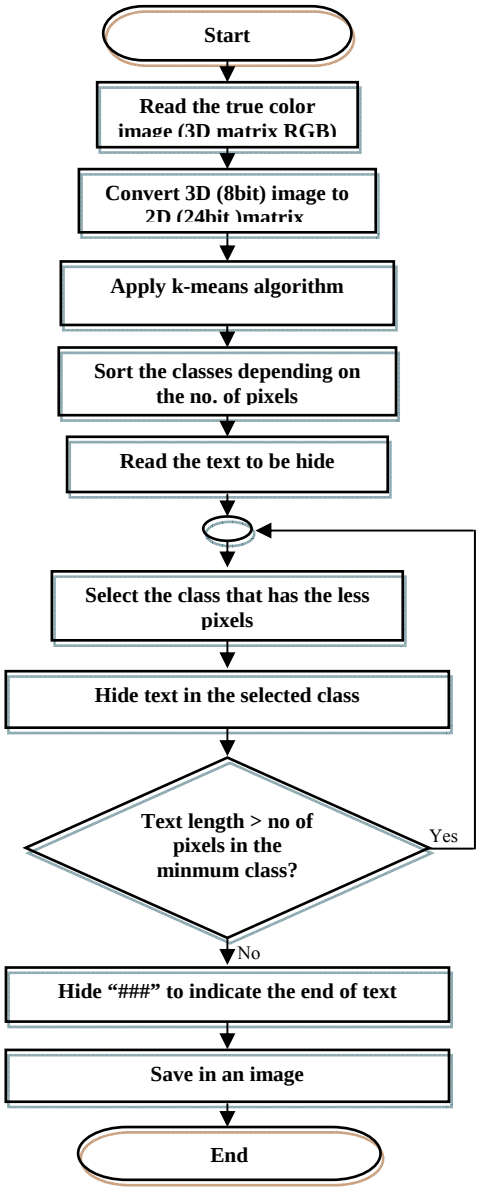


Figure (3), shows the flow chart of the hiding stage in the proposed method

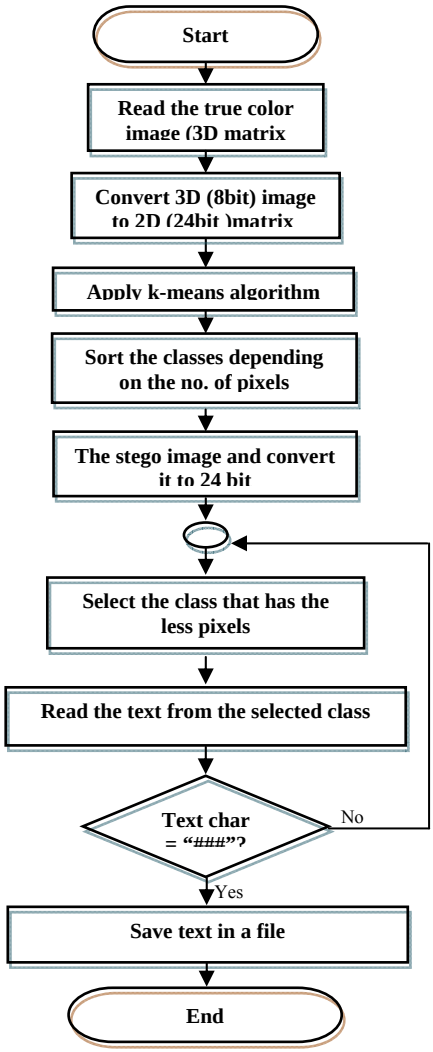


Figure (4), shows the flow chart of the unhiding stage in the proposed method

V. THE RESULTS AND CONCLUSIONS

After applying the proposed algorithm on a number of color images with increasing the number of classes we calculate the correlation factor and the Signal to Noise Ratio between the input image and the resulting images, as shown in the results listed below:

http://sites.google.com/site/ijcsis/
ISSN 1947-5500

1- Figure (5-a) shows the original image, (5-b) resultant image after hiding, the table on the right represented the exchanged classes by text showing the number of pixels that is changed. The number of classes is 17 classes.

The original image

The stego image

CLASS NO.	NUMBER OF PIXELS	CHANGED CLASS
Class 1	262	yes
Class 2	342	yes
Class 3	1450	yes
Class 4	4473	No
Class 5	6535	No
Class 6	6808	No
Class 7	7269	No
Class 8	9840	No
Class 9	12261	No
Class 10	14954	No
Class 11	19809	No
Class 12	20704	No
Class 13	21162	No
Class 14	21666	No
Class 15	23107	No
Class 16	34403	No
Class 17	36155	No

The changed classes by text is mensioned by "yes"

2- Figure (6-a) shows the original image, (6-b) resultant image after hiding, the table on the right represented the exchanged classes by text showing the number of pixels that is changed. The number of classes is 33 classes.

The original image

The stego image

CLASS NO.	NO. OF PIXELS	CHANGED CLASS
Class 1	123	yes
Class 2	124	yes
Class 3	160	yes
Class 4	292	yes
Class 5	432	yes
Class 6	553	yes
Class 7	729	yes
Class 8	842	yes
Class 9	860	yes
Class 10	952	No
Class 11	2168	No
Class 12	2429	No
Class 13	2554	No
Class 14	2698	No
Class 15	2805	No
Class 16	3067	No
Class 17	3192	No
Class 18	3413	No
Class 19	3951	No
Class 20	3969	No
Class 21	3999	No
Class 22	4160	No
Class 23	5229	No
Class 24	6888	No
Class 25	7180	No
Class 26	9253	No
Class 27	12654	No
Class 28	13602	No
Class 29	22283	No
Class 30	26232	No
Class 31	33920	No
Class 32	34170	No
Class 33	38917	No

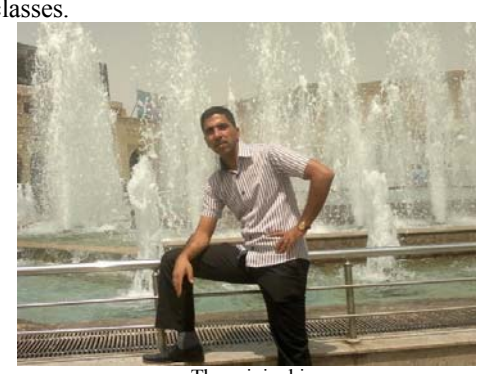
The changed classes by text is mensioned by "yes"

3- Figure (7-a) shows the original image, (7-b) resultant image after hiding, the table on the right represented the exchanged classes by text showing the number of pixels that is changed. The number of classes is 49 classes.

The original image

The stego image

4- Figure (8-a) shows the original image, (8-b) resultant image after hiding, the table on the right represented the exchanged classes by text showing the number of pixels that is changed. The number of classes is 65 classes.



The original image

The stego image

http://sites.google.com/site/ijcsis/
ISSN 1947-5500

As shown in Table (1) that is by increasing the image dimensions the affect of deleting some of the classes are decreased despite the increase in the number of deleted classes, which represents the unimportant features in the images. So, according to this property it is an applicable to use in security applications and sending data on networks.

Table (1): the application of the proposed method on samples of images showing the SNR, PSNR and correlation factor between the original image and the stego image.

Image size	No. of classes	Deleted classes	SNR	PSNR	CORRELATION FACTOR
402×600	17	3	29.9954	60.5275	0.9970
402×600	33	9	29.8022	61.3324	0.9969
402×600	49	15	29.7614	61.2453	0.9969
402×600	65	23	29.6911	61.56227	0.9969
423×600	17	3	34.6233	54.1512	0.9981
423×600	33	9	34.3646	53.915	0.998
423×600	49	15	34.2852	54.0448	0.9979
423×600	65	23	34.75699	54.3150	0.9982
360×638	17	3	33.7461	54.1927	0.9983
360×638	33	9	34.2224	54.2103	0.9985
360×638	49	15	34.1456	54.2042	0.9984
360×638	65	23	33.8570	54.1999	0.9983
393×548	17	3	36.5315	54.8335	0.9982
393×548	33	9	36.4107	54.7552	0.9982
393×548	49	15	36.5128	54.9069	0.9982
393×548	65	23	36.2120	54.7276	0.9981
458×601	17	3	31.7749	53.3484	0.9971
458×601	33	9	31.7027	53.3484	0.9970
458×601	49	15	31.6642	53.3484	0.997
458×601	65	23	31.5652	53.3484	0.9969
589×394	17	3	22.1163	51.2240	0.9809
589×394	33	9	22.1196	51.5429	0.9809
589×394	49	15	22.1144	51.1647	0.9809
589×394	65	23	22.1111	50.9754	0.9808

.

VI. REFERENCES

[1] Siper Alan, Farley Roger and Lombardo Craig, (2005), "The Rise of Steganography", Proceedings of Student/Faculty Research Day, CSIS, Pace University.

[2] Raphael A. Joseph, Sundaram V., A.Joseph, (2011), "Cryptography and Steganography – A Survey", Int. J. Comp. Tech. Appl., Vol 2 (3), 626-630

[3] Gonzalez, R. C. And Woods, R. E., (2008), "Digital Image Processing", Prentice Hall, Inc., 4th edition.

[4] Umbaugh, Scott E., (1998), "Computer Vision And Image Processing", Prentice Hall PTR, USA.

[5] Mumtaz K. and K. Duraiswamy , (2010), "A Novel Density Based Improved K-Means Clustering Algorithm", International Journal on Computer Science and Engineering, India, Vol. 02, No. 02, 213-218.

[6] Ravichandran K.S. And Ananthi B., (2009), "Color skin segmentation using k-means cluster", International Journal of Computational and Applied Mathematics, india volume 4 number 2 pp. 153–157.

AUTHOR PROFILE

Mrs. Shahd A. R. Hasso (M Sc.) is currently a lecturer at Mosul University/ College of Computer Science and Mathematics/ Computer Science Department. She received B.Sc. degree in Computer Science from University of Mosul in 1998 and M.Sc. degree from University of Mosul in 2003. Her research interests and activity are in data security, data strutures, network security, information hiding. Now, she teaches data security undergraduate students

VII. Appendix I: Table shows the number of pixels in each class after k-means of (17, 33, 49, 65) class

Class No.	IMAGE1 classes				IMAGE2 classes				IMAGE3 classes			
	17	33	49	65	17	33	49	65	17	33	49	65
1	262	77	23	26	333	123	58	46	3358	1411	981	563
2	342	78	45	27	1014	124	61	48	3387	1432	990	579
3	1450	85	49	33	1568	160	73	51	3610	1503	990	777
4	4473	99	50	35	1993	292	87	51	3918	1559	1002	791
5	6535	160	51	36	4747	432	100	61	4568	1613	1015	805
6	6808	298	53	37	5803	553	132	74	4723	1632	1039	808
7	7269	604	59	39	5921	729	188	76	4915	1787	1054	811
8	9840	825	70	42	7194	842	257	104	6484	1828	1072	814
9	12261	2156	130	50	8665	860	358	156	7453	2020	1110	821
10	14954	2214	203	52	8893	952	469	182	8812	2104	1119	824
11	19809	2634	427	54	13923	2168	470	215	13525	2185	1186	838
12	20704	4196	624	89	16747	2429	498	250	13765	2475	1207	846
13	21162	4309	828	122	17472	2554	552	308	18805	2604	1344	869
14	21666	4349	900	217	31403	2698	559	351	21241	2831	1380	889
15	23107	4648	1697	311	34629	2805	617	352	21723	3162	1471	894
16	34403	4843	2176	488	45964	3067	726	377	33421	3297	1482	900
17	36155	5115	2211	566	47531	3192	1469	410	55972	3349	1496	955
18		5868	2670	600		3413	1509	414		4254	1649	1004
19		8385	2698	1011		3951	1515	450		4558	1715	1051
20		9657	2882	1084		3969	1527	501		4697	1740	1065
21		9909	2935	1445		3999	1557	590		4768	1762	1097
22		9958	3117	1766		4160	1641	818		6919	1897	1101
23		10562	3128	1824		5229	1721	934		7575	1971	1117
24		10578	3392	1934		6888	1770	1171		7910	2073	1208
25		11040	3775	2014		7180	1820	1196		9426	2163	1301
26		11571	4058	2122		9253	1993	1206		9657	2223	1301
27		11640	4209	2167		12654	2126	1220		11542	2274	1347
28		12491	5889	2195		13602	2181	1230		11905	2571	1361
29		13947	6308	2287		22283	2282	1252		11919	2696	1444
30		13976	6453	2419		26232	2622	1305		12492	2760	1461
31		19232	6644	2485		33920	2808	1366		15860	3502	1603
32		21377	6928	2498		34170	2889	1368		32557	4052	1648
33		24319	7013	2605		38917	3026	1426		36849	4291	1705
34			7211	3080			3353	1439			4484	1742
35			7276	3265			3387	1491			4673	1763
36			7314	3494			4711	1525			5352	1810
37			7423	3997			5133	1553			6089	1828
38			7885	4205			5799	1626			6714	1910
39			7942	4408			7536	1664			7431	2099
40			8069	4939			7574	2038			7551	2166
41			8479	4977			8122	2067			8060	2486
42			8930	5403			11108	2083			8237	2488
43			9160	5412			12903	2197			8253	2816
44			9249	5435			17053	2273			8320	3505
45			10472	5499			20448	2589			8714	3530
46			11091	5637			23523	2650			11969	3605
47			14412	5662			26751	3098			15909	3785
48			16535	5790			27736	3185			25694	3839
49			18048	5793			29002	4031			32953	4580
50				5808				4031				4984
51				5817				4199				5081
52				5863				5636				5511
53				6024				6290				5916
54				6180				6335				6007
55				6255				6466				6154
56				6465				7664				6338
57				7015				11546				6407
58				7156				11790				6474
59				7425				14117				6760
60				7430				14811				6860
61				8596				17025				9027
62				10636				21181				9492
63				12934				21867				16357
64				13202				22123				21969
65				14718				23651				27793

http://sites.google.com/site/ijcsis/
ISSN 1947-5500

Class No.	IMAGE4 classes				IMAGE5 classes				IMAGE6 classes			
	17	33	49	65	17	33	49	65	17	33	49	65
1	252	64	38	24	1459	408	232	182	1177	27	219	170
2	1597	75	41	29	3979	511	290	205	1207	35	227	177
3	2422	82	42	31	8008	627	372	240	2618	49	236	183
4	3287	122	44	31	8138	954	375	252	3174	51	258	187
5	4942	318	48	32	9026	1879	445	284	3777	52	308	189
6	5826	732	50	32	9163	3101	626	296	4475	67	327	195
7	5844	967	80	34	9721	3279	985	340	5599	74	341	209
8	6061	1283	220	39	10955	3665	1373	381	8390	124	356	215
9	7273	1717	578	63	14284	3894	1987	695	17947	178	419	233
10	7563	1963	708	86	14833	4346	2308	929	18030	1988	443	242
11	8320	2297	865	249	17025	4348	2561	1256	19768	2139	668	242
12	9157	2354	929	491	17590	4708	2598	1484	19925	2242	737	274
13	13444	2880	1110	544	18526	4838	2653	1625	22637	2486	1197	375
14	19261	3472	1199	554	19193	5111	2658	1725	23313	2615	1242	417
15	29912	3509	1225	675	24583	5171	2665	1754	24820	3127	1393	510
16	30602	3601	1256	714	34919	5680	2782	1827	26385	3379	1408	718
17	59601	3611	1340	729	53856	5961	2873	2046	28824	5789	1507	746
18		3640	1920	829		7277	2972	2059		9799	1655	841
19		3750	1924	888		7575	3069	2060		10380	1664	957
20		4029	2143	963		8084	3114	2089		10875	1778	1026
21		4100	2244	1000		8425	3350	2097		11250	1887	1040
22		4357	2248	1021		8512	3399	2132		11278	2023	1105
23		4690	2278	1195		8798	3430	2173		11482	2130	1126
24		4966	2279	1324		8854	3823	2271		11577	2745	1226
25		6460	2316	1336		9688	4082	2313		11959	3984	1237
26		8016	2357	1341		10121	4388	2437		12046	5584	1335
27		9099	2374	1394		10280	5169	2453		12739	6631	1362
28		11252	2381	1559		10835	5246	2505		12797	7130	1539
29		15989	2382	1681		12535	5271	2713		14186	7156	1618
30		16970	2479	1755		14498	5300	2843		14440	7293	1620
31		20882	2671	1763		23173	5541	2876		15327	7420	1907
32		28897	2835	1770		24375	5952	2967		15710	7458	1947
33		39220	3124	1779		43747	6003	2979		15803	7648	2926
34			3149	1791			6080	3038			7680	4438
35			3179	1838			6282	3173			7701	4572
36			3209	1862			6343	3536			7787	4717
37			3554	1878			6395	3717			7877	4843
38			4256	1906			6411	3813			7998	5109
39			5521	1930			6440	4117			8657	5163
40			5597	1939			6482	4162			8658	5271
41			6559	1988			7145	4172			8690	5528
42			7169	2014			7460	4187			8863	5571
43			9339	2071			8816	4222			9234	5604
44			12101	2111			8832	4290			9522	5791
45			12265	2181			13553	4612			10118	5838
46			15944	2272			15749	4975			10692	5854
47			16061	2437			18935	5004			10707	5897
48			27381	2468			24052	5041			10963	6024
49			32352	2754			28391	5066			11447	6056
50				3027				5085				6178
51				3741				5098				6242
52				4126				5112				6334
53				4188				5130				6470
54				4377				5146				6779
55				4525				5349				6794
56				6094				5944				6803
57				8273				6739				6864
58				8556				7285				6877
59				9696				9262				7385
60				9891				10228				7460
61				12179				11443				8220
62				13681				11786				8482
63				18836				15713				8618
64				19457				21875				8925
65				25322				22450				9265

http://sites.google.com/site/ijcsis/
ISSN 1947-5500

ANALYSIS OF EXAMINATION RESULTS DATA USING VARIOUS MINING TECHNIQUES

DEVENDRA SINGH RAJPOOT
Ph.D. Scholor , UIT, RGPV,Bhoapl (M.P.)
dsrphd@yahoo.com

Dr. Kanak Saxena
Professor & Head,
Computer Applications
SATI, Vidisha (M.P.)
Kanak.saxena@gmail.com

Dr. Anubhuti Khare
Associate Professor,
DoEC, UIT,
RGPV,Bhopal (M.P.)
anubhutikhare@gmail.com

ABSTRACT

The paper comprises of various pattern mining techniques from data mining such as statistical techniques, classification and clustering. The domain we have chosen is the university domain for the above entitled thesis. The objective for choosing a university domain is, as educational data mining is an emerging discipline concern with the developing method for the exploring the unique types of data that come from the educational context. Due to an increasing number of institutions and students' technical educational institutions becoming increasingly oriented to performance and their measurement and an accordingly setting goals and developing strategies for their achievements [02]. This already happens in Europe in Croatia, USA [01] but still lacking in India. The pattern extracted after applying mining techniques, clearly shows the impact of subject contents in the students' career with the variations in the examination policy.

INTRODUCTION:

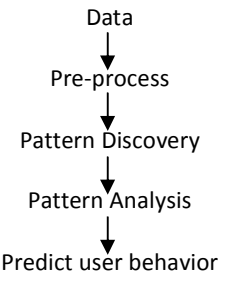
In our mining system the data preprocess is the phase where data cleaned from noise by overcoming the difficulties of recognizing students, semester, branch in order to be used as input to the next phase of pattern discovery. In the pattern mining phase various mining algorithms are incorporated into the system to mine different types of pattern. In the pattern analysis phase the mined patterns which in great number to be evaluated. Mining system is classified and explained.
Commonly a mining system introduces three parts:

(i) Data Preprocessing

(ii) Pattern Discovery

(iii) Pattern Analysis

General Mechanism



DATA DESCRIPTION:

There are about millions of data on students who belongs to various courses, years, semesters etc. Among which we have taken a sample of approx 2 lacs data, When we applied various analytical techniques we found the results of the analysis takes very long time and every time we have to pre-

process the data. Thus for simplicity we have taken a particular semester and a specific range of year from 2004 to 2008 with only one course. The sample data comes out to be near about 16574. For the complete analysis the data are chosen from the university which consists of total attributes 154. Applying the mining algorithms on the complete data the problems of execution due to the constraints of computer system exist. Thus we reduce our data set with approx 16574. No doubt the system accumulates vast amount information which is very valuable for analyzing the student behavior and could create valuable information to the educational system but as discussed earlier, for mining the entire data would not be possible. Hence the data which consider for the valuation is consisting of Engineering III Semester (All disciplines) since the year 2004 to 2008. The interest for performance indicators in the technical education has become extremely high as the reason for this lies in the relevant political and social changes in the recent years [03,04,05,06,07,08,09,10].

WORK DONE:

Data mining is the process of efficient discovery of non-obvious valuable pattern from a large collection of data [11]. To comprehend better the student's behavior, statistical data processing will be performed. In the first segment, graphs will be used to present the basic information on the structure of the student's data and second segment the analysis will be carried out by using various regression techniques.

For this work we use weka 3.6.2 because of its important characteristics [12]:

(i) Free Software System which is implemented in the Java interface.
(ii) Open source software that provides a collection of machine learning and data mining algorithms.
(iii) The algorithms and routines can be modified using the same programming language.

Course	Intake
BE (All discipline)	64430
B. Pharm.	5880
MCA	5980
B. Arch.	300

Table 1. Shows total intake of students of Technical University in the year 2008.

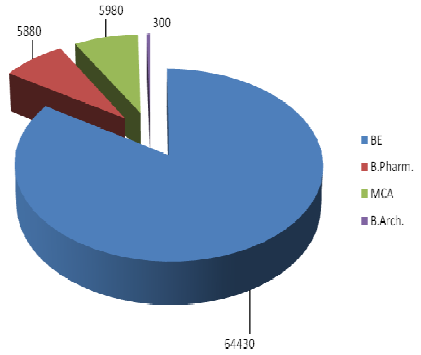


Figure 1.Shows total intake of students of Technical University in the year 2008 with the help of pie chart.

PROPOSED METHOD:

With the increase in demand of technology interest towards technical field is increasing day by day due to which students are taking admission in engineering. As compared to other courses job opportunities are more in the engineering field. The above figure no.1 shows the number of students took admission in engineering for which it is clearly understood that interest of students in engineering is more compared to other courses. B Pharmacy is less in demand due to less number of colleges, limited seats and less job opportunities in this field. Admission in MCA is less because now a day's students prefers to do other courses such as B.Tech. and M.Tech. after bachelor degree of engineering due to number of seats increase. Least admissions are in B. Arch because students interested in this field choose civil engineering as their subject, so admissions in this field are less.

http://sites.google.com/site/ijcsis/
ISSN 1947-5500

Exam_Yr	Std_appear	Std_pass	Result %	Overall Result %
2004	7559	2840	37.57	43.64
2005	8148	4130	50.68	52.35
2006	9484	3992	42.09	49.78
2007	15944	6473	40.59	43.51
2008	17731	10475	59.07	52.18

Table 2.: Shows number of students in Engineering and their result from 2004 to 2008.

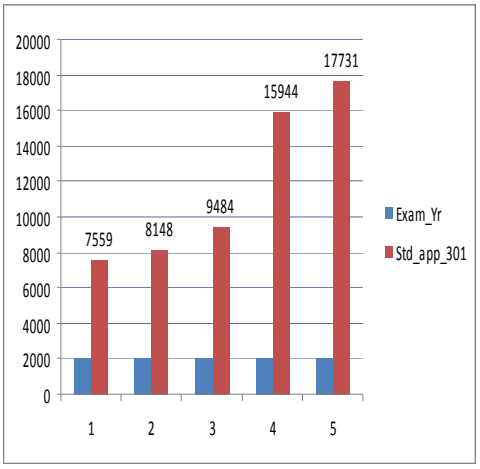


Figure 2. Shows number of students in Engineering since 2004 to 2008.

RESULT DISCUSSIONS:

Due to increase in engineering colleges as well as an increase in intake in the state, Number of students appearing in exams are also increasing. As per the table no.2. Number of students appeared and the number of students passed in these exams have also shown the trend in decreased of overall results with every year. After analysis we found that failure rate is more than pass rate in more students are failing to clear the subject of Mathematics-III. In year 2004, 7559 students were appeared in the examination and 2840 are successful to clear and 4719 students are failing in Mathematics-III, like while in year 2005, 4130 students pass out of 8148 and 4318 students are failing, year 2006, 3992 students pass out of 9484, year 2007, 6473 students pass out of 15944 and year 2008, 10475 students pass out of 17731.

For this we have used the classification techniques a classifier is a mapping from X to a discrete set of labels Y [13]. These analyses predict the class label which is based on supervised learning and provides a collection of labeled i.e. Pre classified pattern. The classification has been used for discovering the students' behavior which similar characteristics and reaction to a specific pedagogical strategies [14], predicting students' performance [15] as well as the relevance of the examination paper in a semester (Regular as well as back papers) involved.

Classification Method	*Mode of Test*	*Correctly Classified Instances*	*Incorrectly Classified Instances*
Decision Table	10 fold	14732	518
	75% splitting	3697	124
	Training set	14768	482
REPtree	10 fold	14570	680
	75% splitting	3657	164
	Training set	14570	680

Table 3. Correctly classified and incorrectly classified instances on different classification methods and mode of the test.

We have performed total 6 classification experiments on the university data, Decision Table & REPtree method with three different Test Mode (10 Cross Fold, 75% split, Full training set). Which is shown in table No.3 and figure No.3.

http://sites.google.com/site/ijcsis/
ISSN 1947-5500

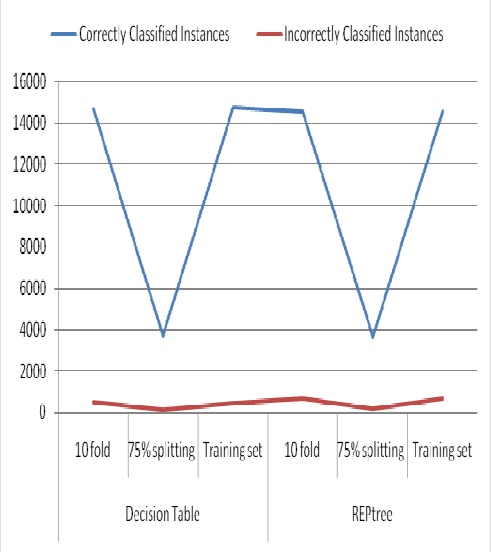


Figure 3. Decision Table & REPtree method with three different Test Mode

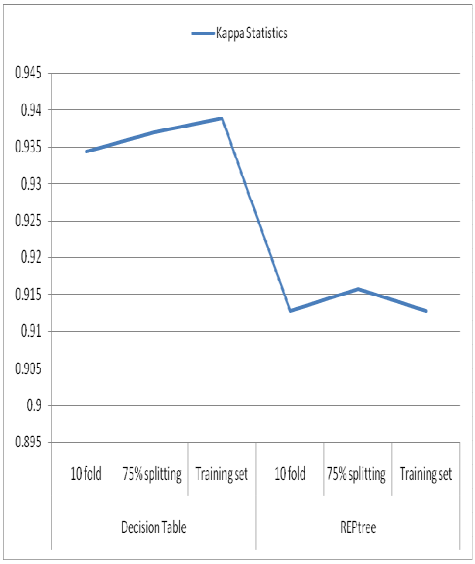


Figure 4. Kappa Statistics on different classification methods and mode of test.

Decision table classification methods classify correctly the highest number of instances 14768, while data size (16000 x 27) is taken as training set. REPtree classification methods classify correctly the lowest number of instances 14570, while data size (16000 x 27) is taken as training set.

Classification Method	***Mode of Test***	***Kappa Statistics***
Decision Table	10 fold	0.9343
	75% splitting	0.9369
	Training set	0.9388
REPtree	10 fold	0.9128
	75% splitting	0.9157
	Training set	0.9128

Table 4. Kappa Statistics on different classification methods and mode of test.

Decision table classification methods calculate the highest kappa statistics 0.9388. Kappa is a measure of agreement normalized for chance agreement.

$$K = \frac{P(A) - P(E)}{1 - P(E)}$$

Where P (A) is the percentage agreement (e.g., Between your classifier and ground truth) and P(E) is the chance agreement.

K=1 indicates perfect agreement,

K=0 indicates chance agreement.

Kappa is a chance-corrected measure of agreement between the classifications and the true classes. It's calculated by taking the agreement expected by chance away from the observed agreement and dividing by the maximum possible agreement. A value greater than 0 means that your classifier is doing better than chance.

Classification Method	Mode of Test	weighted average TP rate	weighted average FP rate	weighted average Precision	weighted average recall	weighted average F-Measure	Time Taken (second)
Decision Table	10 fold	0.966	0.025	0.968	0.966	0.96	27.98
	75% splitting	0.968	0.024	0.969	0.968	0.962	29.06
	Training set	0.968	0.023	0.97	0.968	0.963	27.77
REPtree	10 fold	0.955	0.033	0.92	0.955	0.937	36.24
	75% splitting	0.957	0.031	0.922	0.957	0.939	38.08
	Training set	0.955	0.033	0.92	0.955	0.937	37.23

Table 5. Classification Factors of Decision Table, REPtree on different test mode

CONCLUSIONS AND FUTURE WORK:

In this work analysis of examination data has been done. Classification of data has been done using Decision table and REPtree and Kappa statistics has played its own role. Work done has been compared with the help of well known tool, which shows good results. In future some more data will be taken to anylaysed results.

REFERENCES:

[01] Al-Hawaj, A. Y., Elali, W., Twizell, E. H. (Ed.) (2008): Higher Education in the Twenty-First Century: Issues and Challenges, Taylor & Francis Group, London

[02] Pausits, A., Pellert, A. (2007): Higher Education Management and Development in Central, Southern and Eastern Europe, WAXMANN Verlag, Munster

[03] GFME (2008): The Global Management Education Landscape: Shaping the future of business schools, Global Foundation for Management Education

[04] McKelvey, M., Holmén, M. (Ed.) (2009): Learning to Compete in European Universities: From Social Institution to Knowledge Business, Edward Elgar Publishing, Inc., Massachusetts

[05] NCVVO (2009):Vodič za provedbu samovrjednovanja u osnovnim školama, Nacionalni centar za vanjsko vrednovanje obrazovanja, Zagreb

[06] Vašiček, V., Budimir, V., Letinić, S. (2007): Pokazatelji uspješnosti u visokom obrazovanju, Privredna kretanja i ekonomska politika, 17 (110): str. 51 - 80.

[07] Orsingher, Ch. (Ed.) (2006): Assessing Quality in European Higher Education Institutions: Dissemination, Methods and Procedures, Physica-Verlag: Springer, Heidelberg

[08] Knust, M., Hanft, A. (Ed.) (2009): Continuing Higher Education and Lifelong Learning: An International Comparative Study on Structures, Organisation and Provisions, Springer Science & Business Media, Heidelberg

[09] Deem, R., Hillyard, S., Reed, M. (2007): Knowledge, Higher Education, and the New Managerialism: The Changing Management of UK Universities, Oxford University Press Inc., New York

[10] Michael, S. O., Kretovics, M. A. (Ed.) (2005): Financing Higher Education in a Global Market, Algora Publishing, New York

[11] Klosgen, W., & Zytkow, J. (2002). Handbook of data mining and knowledge discovery. New York: Oxford University Press.

[12] Witten, I. H., & Frank, E. (2005). Data mining: Practical machine learning tools and techniques. Morgan Kaufman.

[13] Duda, R. O., Hart, P. E., & Stork, D. G. (2000). Pattern classification. Wiley Interscience.

[14] Chen, G., Liu, C., Ou, K., & Liu, B. (2000). Discovering decision knowledge from web log portfolio for managing classroom processes by applying decision tree and data cube technology. Journal of Educational Computing Research, 23(3), 305–332.

[15] Minaei-Bidgoli, B., & Punch, W. (2003). Using genetic algorithms for data mining optimization in an educational web-based system. In Genetic and evolutionary computation conference, Chicago, USA (pp. 2252–2263).

IJCSIS REVIEWERS' LIST

Assist Prof (Dr.) M. Emre Celebi, Louisiana State University in Shreveport, USA
Dr. Lam Hong Lee, Universiti Tunku Abdul Rahman, Malaysia
Dr. Shimon K. Modi, Director of Research BSPA Labs, Purdue University, USA
Dr. Jianguo Ding, Norwegian University of Science and Technology (NTNU), Norway
Assoc. Prof. N. Jaisankar, VIT University, Vellore,Tamilnadu, India
Dr. Amogh Kavimandan, The Mathworks Inc., USA
Dr. Ramasamy Mariappan, Vinayaka Missions University, India
Dr. Yong Li, School of Electronic and Information Engineering, Beijing Jiaotong University, P.R. China
Assist. Prof. Sugam Sharma, NIET, India / Iowa State University, USA
Dr. Jorge A. Ruiz-Vanoye, Universidad Autónoma del Estado de Morelos, Mexico
Dr. Neeraj Kumar, SMVD University, Katra (J&K), India
Dr Genge Bela, "Petru Maior" University of Targu Mures, Romania
Dr. Junjie Peng, Shanghai University, P. R. China
Dr. Ilhem LENGLIZ, HANA Group - CRISTAL Laboratory, Tunisia
Prof. Dr. Durgesh Kumar Mishra, Acropolis Institute of Technology and Research, Indore, MP, India
Jorge L. Hernández-Ardieta, University Carlos III of Madrid, Spain
Prof. Dr.C.Suresh Gnana Dhas, Anna University, India
Mrs Li Fang, Nanyang Technological University, Singapore
Prof. Pijush Biswas, RCC Institute of Information Technology, India
Dr. Siddhivinayak Kulkarni, University of Ballarat, Ballarat, Victoria, Australia
Dr. A. Arul Lawrence, Royal College of Engineering & Technology, India
Mr. Wongyos Keardsri, Chulalongkorn University, Bangkok, Thailand
Mr. Somesh Kumar Dewangan, CSVTU Bhilai (C.G.)/ Dimat Raipur, India
Mr. Hayder N. Jasem, University Putra Malaysia, Malaysia
Mr. A.V.Senthil Kumar, C. M. S. College of Science and Commerce, India
Mr. R. S. Karthik, C. M. S. College of Science and Commerce, India
Mr. P. Vasant, University Technology Petronas, Malaysia
Mr. Wong Kok Seng, Soongsil University, Seoul, South Korea
Mr. Praveen Ranjan Srivastava, BITS PILANI, India
Mr. Kong Sang Kelvin, Leong, The Hong Kong Polytechnic University, Hong Kong
Mr. Mohd Nazri Ismail, Universiti Kuala Lumpur, Malaysia
Dr. Rami J. Matarneh, Al-isra Private University, Amman, Jordan
Dr Ojesanmi Olusegun Ayodeji, Ajayi Crowther University, Oyo, Nigeria
Dr. Riktesh Srivastava, Skyline University, UAE
Dr. Oras F. Baker, UCSI University - Kuala Lumpur, Malaysia
Dr. Ahmed S. Ghiduk, Faculty of Science, Beni-Suef University, Egypt
and Department of Computer science, Taif University, Saudi Arabia
Mr. Tirthankar Gayen, IIT Kharagpur, India
Ms. Huei-Ru Tseng, National Chiao Tung University, Taiwan

CALL FOR PAPERS
International Journal of Computer Science and Information Security
January - December
IJCSIS 2012
ISSN: 1947-5500
http://sites.google.com/site/ijcsis/

International Journal Computer Science and Information Security, IJCSIS, is the premier scholarly venue in the areas of computer science and security issues. IJCSIS 2011 will provide a high profile, leading edge platform for researchers and engineers alike to publish state-of-the-art research in the respective fields of information technology and communication security. The journal will feature a diverse mixture of publication articles including core and applied computer science related topics.

Authors are solicited to contribute to the special issue by submitting articles that illustrate research results, projects, surveying works and industrial experiences that describe significant advances in the following areas, but are not limited to. Submissions may span a broad range of topics, e.g.:

Track A: Security

Access control, Anonymity, Audit and audit reduction & Authentication and authorization, Applied cryptography, Cryptanalysis, Digital Signatures, Biometric security, Boundary control devices, Certification and accreditation, Cross-layer design for security, Security & Network Management, Data and system integrity, Database security, Defensive information warfare, Denial of service protection, Intrusion Detection, Anti-malware, Distributed systems security, Electronic commerce, E-mail security, Spam, Phishing, E-mail fraud, Virus, worms, Trojan Protection, Grid security, Information hiding and watermarking & Information survivability, Insider threat protection, Integrity
Intellectual property protection, Internet/Intranet Security, Key management and key recovery, Language-based security, Mobile and wireless security, Mobile, Ad Hoc and Sensor Network Security, Monitoring and surveillance, Multimedia security ,Operating system security, Peer-to-peer security, Performance Evaluations of Protocols & Security Application, Privacy and data protection, Product evaluation criteria and compliance, Risk evaluation and security certification, Risk/vulnerability assessment, Security & Network Management, Security Models & protocols, Security threats & countermeasures (DDoS, MiM, Session Hijacking, Replay attack etc,), Trusted computing, Ubiquitous Computing Security, Virtualization security, VoIP security, Web 2.0 security, Submission Procedures, Active Defense Systems, Adaptive Defense Systems, Benchmark, Analysis and Evaluation of Security Systems, Distributed Access Control and Trust Management, Distributed Attack Systems and Mechanisms, Distributed Intrusion Detection/Prevention Systems, Denial-of-Service Attacks and Countermeasures, High Performance Security Systems, Identity Management and Authentication, Implementation, Deployment and Management of Security Systems, Intelligent Defense Systems, Internet and Network Forensics, Large-scale Attacks and Defense, RFID Security and Privacy, Security Architectures in Distributed Network Systems, Security for Critical Infrastructures, Security for P2P systems and Grid Systems, Security in E-Commerce, Security and Privacy in Wireless Networks, Secure Mobile Agents and Mobile Code, Security Protocols, Security Simulation and Tools, Security Theory and Tools, Standards and Assurance Methods, Trusted Computing, Viruses, Worms, and Other Malicious Code, World Wide Web Security, Novel and emerging secure architecture, Study of attack strategies, attack modeling, Case studies and analysis of actual attacks, Continuity of Operations during an attack, Key management, Trust management, Intrusion detection techniques, Intrusion response, alarm management, and correlation analysis, Study of tradeoffs between security and system performance, Intrusion tolerance systems, Secure protocols, Security in wireless networks (e.g. mesh networks, sensor networks, etc.), Cryptography and Secure Communications, Computer Forensics, Recovery and Healing, Security Visualization, Formal Methods in Security, Principles for Designing a Secure Computing System, Autonomic Security, Internet Security, Security in Health Care Systems, Security Solutions Using Reconfigurable Computing, Adaptive and Intelligent Defense Systems, Authentication and Access control, Denial of service attacks and countermeasures, Identity, Route and

Location Anonymity schemes, Intrusion detection and prevention techniques, Cryptography, encryption algorithms and Key management schemes, Secure routing schemes, Secure neighbor discovery and localization, Trust establishment and maintenance, Confidentiality and data integrity, Security architectures, deployments and solutions, Emerging threats to cloud-based services, Security model for new services, Cloud-aware web service security, Information hiding in Cloud Computing, Securing distributed data storage in cloud, Security, privacy and trust in mobile computing systems and applications, **Middleware security & Security features:** middleware software is an asset on
its own and has to be protected, interaction between security-specific and other middleware features, e.g., context-awareness, **Middleware-level security monitoring and measurement:** metrics and mechanisms for quantification and evaluation of security enforced by the middleware, **Security co-design:** trade-off and co-design between application-based and middleware-based security, **Policy-based management:** innovative support for policy-based definition and enforcement of security concerns, **Identification and authentication mechanisms:** Means to capture application specific constraints in defining and enforcing access control rules, **Middleware-oriented security patterns:** identification of patterns for sound, reusable security, **Security in aspect-based middleware:** mechanisms for isolating and enforcing security aspects, **Security in agent-based platforms:** protection for mobile code and platforms, Smart Devices: Biometrics, National ID cards, Embedded Systems Security and TPMs, RFID Systems Security, Smart Card Security, Pervasive Systems: Digital Rights Management (DRM) in pervasive environments, Intrusion Detection and Information Filtering, Localization Systems Security (Tracking of People and Goods), Mobile Commerce Security, Privacy Enhancing Technologies, Security Protocols (for Identification and Authentication, Confidentiality and Privacy, and Integrity), Ubiquitous Networks: Ad Hoc Networks Security, Delay-Tolerant Network Security, Domestic Network Security, Peer-to-Peer Networks Security, Security Issues in Mobile and Ubiquitous Networks, Security of GSM/GPRS/UMTS Systems, Sensor Networks Security, Vehicular Network Security, Wireless Communication Security: Bluetooth, NFC, WiFi, WiMAX, WiMedia, others

This Track will emphasize the design, implementation, management and applications of computer communications, networks and services. Topics of mostly theoretical nature are also welcome, provided there is clear practical potential in applying the results of such work.

Track B: Computer Science

Broadband wireless technologies: LTE, WiMAX, WiRAN, HSDPA, HSUPA, Resource allocation and interference management, Quality of service and scheduling methods, Capacity planning and dimensioning, Cross-layer design and Physical layer based issue, Interworking architecture and interoperability, Relay assisted and cooperative communications, Location and provisioning and mobility management, Call admission and flow/congestion control, Performance optimization, Channel capacity modeling and analysis, Middleware Issues: Event-based, publish/subscribe, and message-oriented middleware, Reconfigurable, adaptable, and reflective middleware approaches, Middleware solutions for reliability, fault tolerance, and quality-of-service, Scalability of middleware, Context-aware middleware, Autonomic and self-managing middleware, Evaluation techniques for middleware solutions, Formal methods and tools for designing, verifying, and evaluating, middleware, Software engineering techniques for middleware, Service oriented middleware, Agent-based middleware, Security middleware, Network Applications: Network-based automation, Cloud applications, Ubiquitous and pervasive applications, Collaborative applications, RFID and sensor network applications, Mobile applications, Smart home applications, Infrastructure monitoring and control applications, Remote health monitoring, GPS and location-based applications, Networked vehicles applications, Alert applications, Embeded Computer System, Advanced Control Systems, and Intelligent Control : Advanced control and measurement, computer and microprocessor-based control, signal processing, estimation and identification techniques, application specific IC's, nonlinear and adaptive control, optimal and robot control, intelligent control, evolutionary computing, and intelligent systems, instrumentation subject to critical conditions, automotive, marine and aero-space control and all other control applications, Intelligent Control System, Wiring/Wireless Sensor, Signal Control System. Sensors, Actuators and Systems Integration : Intelligent sensors and actuators, multisensor fusion, sensor array and multi-channel processing, micro/nano technology, microsensors and microactuators, instrumentation electronics, MEMS and system integration, wireless sensor, Network Sensor, Hybrid

Sensor, Distributed Sensor Networks. Signal and Image Processing : Digital signal processing theory, methods, DSP implementation, speech processing, image and multidimensional signal processing, Image analysis and processing, Image and Multimedia applications, Real-time multimedia signal processing, Computer vision, Emerging signal processing areas, Remote Sensing, Signal processing in education. Industrial Informatics: Industrial applications of neural networks, fuzzy algorithms, Neuro-Fuzzy application, bioInformatics, real-time computer control, real-time information systems, human-machine interfaces, CAD/CAM/CAT/CIM, virtual reality, industrial communications, flexible manufacturing systems, industrial automated process, Data Storage Management, Harddisk control, Supply Chain Management, Logistics applications, Power plant automation, Drives automation. Information Technology, Management of Information System : Management information systems, Information Management, Nursing information management, Information System, Information Technology and their application, Data retrieval, Data Base Management, Decision analysis methods, Information processing, Operations research, E-Business, E-Commerce, E-Government, Computer Business, Security and risk management, Medical imaging, Biotechnology, Bio-Medicine, Computer-based information systems in health care, Changing Access to Patient Information, Healthcare Management Information Technology. Communication/Computer Network, Transportation Application : On-board diagnostics, Active safety systems, Communication systems, Wireless technology, Communication application, Navigation and Guidance, Vision-based applications, Speech interface, Sensor fusion, Networking theory and technologies, Transportation information, Autonomous vehicle, Vehicle application of affective computing, Advance Computing technology and their application : Broadband and intelligent networks, Data Mining, Data fusion, Computational intelligence, Information and data security, Information indexing and retrieval, Information processing, Information systems and applications, Internet applications and performances, Knowledge based systems, Knowledge management, Software Engineering, Decision making, Mobile networks and services, Network management and services, Neural Network, Fuzzy logics, Neuro-Fuzzy, Expert approaches, Innovation Technology and Management : Innovation and product development, Emerging advances in business and its applications, Creativity in Internet management and retailing, B2B and B2C management, Electronic transceiver device for Retail Marketing Industries, Facilities planning and management, Innovative pervasive computing applications, Programming paradigms for pervasive systems, Software evolution and maintenance in pervasive systems, Middleware services and agent technologies, Adaptive, autonomic and context-aware computing, Mobile/Wireless computing systems and services in pervasive computing, Energy-efficient and green pervasive computing, Communication architectures for pervasive computing, Ad hoc networks for pervasive communications, Pervasive opportunistic communications and applications, Enabling technologies for pervasive systems (e.g., wireless BAN, PAN), Positioning and tracking technologies, Sensors and RFID in pervasive systems, Multimodal sensing and context for pervasive applications, Pervasive sensing, perception and semantic interpretation, Smart devices and intelligent environments, Trust, security and privacy issues in pervasive systems, User interfaces and interaction models, Virtual immersive communications, Wearable computers, Standards and interfaces for pervasive computing environments, Social and economic models for pervasive systems, Active and Programmable Networks, Ad Hoc & Sensor Network, Congestion and/or Flow Control, Content Distribution, Grid Networking, High-speed Network Architectures, Internet Services and Applications, Optical Networks, Mobile and Wireless Networks, Network Modeling and Simulation, Multicast, Multimedia Communications, Network Control and Management, Network Protocols, Network Performance, Network Measurement, Peer to Peer and Overlay Networks, Quality of Service and Quality of Experience, Ubiquitous Networks, Crosscutting Themes – Internet Technologies, Infrastructure, Services and Applications; Open Source Tools, Open Models and Architectures; Security, Privacy and Trust; Navigation Systems, Location Based Services; Social Networks and Online Communities; ICT Convergence, Digital Economy and Digital Divide, Neural Networks, Pattern Recognition, Computer Vision, Advanced Computing Architectures and New Programming Models, Visualization and Virtual Reality as Applied to Computational Science, Computer Architecture and Embedded Systems, Technology in Education, Theoretical Computer Science, Computing Ethics, Computing Practices & Applications

Authors are invited to submit papers through e-mail ijcsiseditor@gmail.com. Submissions must be original and should not have been published previously or be under consideration for publication while being evaluated by IJCSIS. Before submission authors should carefully read over the journal's Author Guidelines, which are located at http://sites.google.com/site/ijcsis/authors-notes .

ISSN 1947 5500
http://sites.google.com/site/ijcsis/